

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
март–апрель 2021 Том 21 № 2 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
March–April 2021 Vol. 21 No 2 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)



КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

COMPUTER SCIENCE

doi: 10.17586/2226-1494-2021-21-2-206-224
УДК 004.852

Инвариантный к линейным конформным преобразованиям алгоритм обнаружения размытого изображения целевого объекта малого размера

Андрей Владимирович Тимофеев[1]✉

[1] ТОО «Эквалайзум», Астана, 010000, Казахстан

[1] timofeev.andrey@gmail.com✉, https://orcid.org/0000-0001-7212-5230

Аннотация

Предмет исследования. Предложен новый, практически эффективный метод обнаружения и оценивания координат центра изображения целевого объекта малого размера на зашумленной сцене. Метод является инвариантным к линейным конформным преобразованиям (поворот, сдвиг и масштаб). **Метод.** Рассмотрен бинарный классификатор, который принимает решение: содержит ли конкретный участок сцены искомое изображение или только фон? В основе предложенного подхода лежит интерактивная процедура поиска экстремума функции, которая аппроксимирует функцию правдоподобия бинарного классификатора. Для реализации поиска экстремума использована двухэтапная процедура на основании метода Нелдера–Мида. С целью обеспечения устойчивости к помехам и линейным конформным преобразованиям, при обучении классификатора применены как специальные методы обучения, так и подход, основанный на использовании ансамбля классификаторов, каждый из которых соответствует определенному масштабу. **Основные результаты.** Создан метод обнаружения размытого изображения объекта малого размера на сцене, зашумленной коррелированной помехой, с одновременным оцениванием координат центра искомого изображения. Метод устойчив к линейным конформным преобразованиям и успешно протестирован как на модельных, так и на реальных изображениях. Робастность метода к коррелированной помехе аддитивного типа, а также устойчивость к линейным конформным преобразованиям подтверждена результатами численного исследования. В рамках предложенного подхода, формально решена задача построения доверительного множества для координат центра искомого изображения, а также численно исследована эффективность полученного решения. Свойства доверительного множества оформлены в виде теоремы. Выполнено сравнение с классическим, корреляционно-экстремальным методом. При необходимости, предложенный метод легко обобщается на случай мультиклассовой постановки задачи. **Практическая значимость.** Метод может быть использован в системах машинного зрения, в том числе для оперативного анализа данных аэрофотосъемки, а также в системах видеоконтроля механического состояния сложного технического оборудования в условиях сильных метеорологических помех.

Ключевые слова
совпадение изображений, машинное обучение, SVM-классификатор, метод Нелдера–Мида

Ссылка для цитирования: Тимофеев А.В. Инвариантный к линейным конформным преобразованиям алгоритм обнаружения размытого изображения целевого объекта малого размера // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21, № 2. С. 206–224. doi: 10.17586/2226-1494-2021-21-2-206-224

Detection of a small target object in blurry images affected by affine distortions

Andrey V. Timofeev[1]✉

[1] LLP EqualiZoom, Astana, 010000, Republic of Kazakhstan

[1] timofeev.andrey@gmail.com✉, https://orcid.org/0000-0001-7212-5230

Abstract

The paper proposes a novel and practically effective method for detecting, classifying, and estimating the coordinates of the image center of a small-size target object on a noisy scene, which is invariant to linear conformal transformations (rotation, shift, and scale). We consider a binary classifier that decides whether a particular part of the scene contains the desired image or only the background. The proposed approach implies an interactive procedure for finding an extremum of a function that approximates the likelihood function of the binary classifier. A two-step procedure based on the Nelder-Meade method is used to implement the extremum search. In order to ensure the robustness to noise and linear conformal transformations, both special training methods and the approach based on using an ensemble of classifiers, each of which corresponds to a certain scale, are applied in training the classifier. The author created a method for detecting a blurred image of a small-sized object in a scene that is distorted by correlated noise and proposes simultaneous estimation of the coordinates of the center of the target image. The method is robust to linear conformal distortions and has been successfully tested both on the artificial model and real images. The results of numerical study confirmed the robustness of the method to correlated noise of additive type and to linear conformal transformations. Within the framework of the proposed approach, the problem of constructing a confidence set for the coordinates of the target image center has been formally solved, and the efficiency of the obtained solution has been numerically investigated. The properties of the confidence set are formalized in the form of a theorem. The work also makes a comparison with the classical correlation-extreme method. If necessary, the proposed method can be easily generalized to the multiclass case. The method can be applied to machine vision systems, including online analysis of aerial survey data and to systems for video monitoring of the mechanical condition of complex technical equipment under conditions of strong meteorological disturbance.

Keywords

image matching, machine learning, SVM-classifier, Nelder-Meade method

For citation: Timofeev A.V. Detection of a small target object in blurry images affected by affine distortions. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2021, vol. 21, no. 2, pp. 206–224 (in Russian). doi: 10.17586/2226-1494-2021-21-2-206-224

Введение

В приложениях технического зрения, например, в задачах оперативного анализа данных аэрофотосъемки и видеоконтроле механического состояния сложного технического оборудования в условиях сильных метеорологических помех, возникает необходимость обнаружения на анализируемой сцене объекта малой формы, с сильно размытым контуром в условиях интенсивной коррелированной помехи. Ввиду размытости контура изображения целевого объекта, а также из-за его сравнительно малой площади, методы детекции целевого объекта, основанные на использовании мелких геометрических особенностей изображения, оказываются практически неэффективными. В частности, для изображений упомянутого типа, такие известные методы как SIFT (Scale-invariant Feature Transform) [1], SURF (Speeded Up Robust Features) [2], ORB (Oriented FAST and Rotated BRIEF) [3], FAST (Fast Affine Template Matching) [4] практически неработоспособны. Все эти методы, в различной форме, используют в своей работе оценки локального градиента контраста, оценки гессиана контраста или подобные численные характеристики локальных областей изображения, которые в дальнейшем объединяются в так называемые дескрипторы (числовые наборы определенной мощности). Далее эти дескрипторы используются для описания множества ключевых точек изображения, описывая его контура, углы и иные ключевые точки. Но для размытых, интенсивно зашумленных изображений качественно рассчитать оценки градиентов, гессианов и иных характеристик изображений, основанных на локальных перепадах яркости, очень сложно, а в ряде случаев и просто невозможно.

В этом состоит основная причина потери работоспособности упомянутой группы методов для размытых целевых изображений малого размера. Для обнаружения и одновременной классификации целевых изображений этого типа логично использовать методы, которые единомоментно исследуют изображения в некоторой, пространственно распределенной области, с размером, сравнимым с радиусом корреляции искомого изображения. В основе работы метода лежит концепция формирования геометрически целостного образа искомого изображения, причем этот образ единовременно собирается из группы пространственно разнесенных точек, в отличие от методов, основанных на исследовании локального контраста, которые сначала формируют множество разрозненных признаков и далее формируют из них образ, который сравнивается с эталоном. К методам такого типа, например, относится классический, корреляционно-экстремальный метод [5–7], а также его современные обобщения [8, 9]. В настоящей работе предложен метод обнаружения и классификации, с одновременным оцениванием центра целевого, размытого изображения небольшого размера на сцене, зашумленной коррелированной помехой. Метод основан на исследовании сцены в областях, размер которых сравним с радиусом корреляции искомого изображения и сохраняет работоспособность в условиях сильной размытости контуров при наличии интенсивного, коррелированного шума. Метод

устойчив к линейным конформным преобразованиям формы искомого объекта (поворот, сдвиг и масштаб) относительно шаблона. Проведено сравнение с классическим, корреляционно-экстремальным методом, а в силу принципа, заложенного в основу предложенного метода, он легко распространяется на мультиклассовый случай.

Идея метода

В данном разделе приведено неформальное описание предложенного подхода для тех читателей, которые не планируют погружение в формальную базу и после ознакомления с настоящим разделом могут продолжить чтение сразу с раздела «Результаты исследования алгоритма». Основная идея предложенного метода основана на введении бинарного классификатора, который принимает решение: содержит ли конкретный участок сцены искомое изображение или в данной точке сцены присутствует только фон? Важную роль играет функция правдоподобия этого классификатора, которая теоретически достигает своего максимума именно в точке местоположения искомого объекта. Это обусловлено логикой обучения классификатора, которое производится по специальной схеме. Для рассматриваемого случая задача бинарной классификации изображения конкретного участка сцены эквивалентна задаче обнаружения (детекции) искомого изображения на сцене. В этом случае «сигнальный» класс представляет собой различные варианты представления искомого изображения, а «фоновый» класс включает в себя изображения фона или фрагментов нецелевых изображений. Важно то, что оценка координат точки максимума функции правдоподобия, де-факто, является оценкой центра искомого изображения. Таким образом, задачи обнаружения и оценивания координат центра искомого изображения инкапсулированы в процедуру бинарной классификации изображения конкретного участка сцены, которая, в виде побочного результата, формирует оценку координат центра изображения. В принципе, предлагаемый алгоритм подразумевает возможность обнаруживать одним и тем же классификатором сразу несколько типов целевых изображений, отличая их и от фона, и друг от друга. В этом случае речь идет о мультиклассовой постановке задачи, которая в данной статье не рассмотрена. Упрощая, можно считать, что предложенный метод является сильным обобщением методов совпадения изображений.

Для аппроксимации функции правдоподобия классификатора используются широко известные методы, например масштабирование Платта. Для интеграционного поиска точки экстремума функции правдоподобия в области, сравнимой с радиусом корреляции искомого изображения, использован практически эффективный метод Нелдера–Мида. С целью повышения вероятности поиска глобального экстремума поисковая процедура построена по двухэтапной схеме:
— на первом этапе используется классификатор с более «широкой» функцией правдоподобия;
— на втором этапе используется классификатор с «узкой» и более «точной» функцией правдоподобия, где в качестве начальной точки используются результаты первого этапа.

Вычислительные эксперименты показали, что предложенный алгоритм робастен к аддитивному коррелированному шуму в следующем смысле: средняя ошибка оценивания центра искомого изображения при увеличении интенсивности шума нарастает сравнительно плавно. Сделана попытка формализовать этот тип робастности. Важная задача — обеспечение сохранения работоспособности алгоритма в условиях линейных конформных преобразований (сдвиг, поворот, масштаб), когда априорно неизвестно: под каким углом будет расположено целевое изображение на сцене и каков будет его масштаб? Решение этой задачи стало центральным моментом настоящей работы и достигнуто композицией двух подходов.

Первый подход: для преодоления неопределенности относительно значения масштаба использовался ансамбль классификаторов, каждый из которых был настроен методами машинного обучения на определенное значение масштаба, а результирующее решение представляло собой взвешенную сумму этих классификаторов. Значения масштаба, на которые настраиваются соответствующие классификаторы ансамбля — последовательность чисел (масштабная сетка) с шагом n % (n больше нуля, конкретное значение этого параметра выбирается специальными методами), внутри интервала, который определяет априорно допустимые вариации масштаба. Вычислительные эксперименты показали, что при шаге масштабной сетки равном 5 %, система гарантировала вероятность правильной классификации (детекции) не хуже, чем 0,99, при величине средней ошибки оценивания координат центра искомого изображения, не превышающей величину 4 %.

Второй подход: преодоление априорной неопределенности относительно угла поворота искомого изображения на сцене. В рамках подхода в обучающие множества, соответствующие каждому элементу масштабной сетки, внедрялись варианты образцов искомого изображения, углы поворота которых представляли собой последовательность значений углов поворота (ротационную сетку) с шагом 5 % внутри интервала $(0,2\pi)$. Этот достаточно простой подход также доказал свою практическую эффективность.

Определения и обозначения

Обозначим символом o — целевой объект. Любому объекту o соответствует счетное множество его возможных изображений $\mathbf{F}^{(o)} = \{\mathbf{f}_i\}$. Изображение конкретного, целевого объекта будем называть объектовым.

В общем случае изображения могут не быть объектовыми и содержать:
— изображения только подстилающей поверхности (фона);
— изображения фрагментов группы целевых или нецелевых объектов.

Такие изображения будем называть **фоновыми**.

Для множества $\mathbf{F}^{(o)}$ определено конечное подмножество $F^{(o)} \subset \mathbf{F}^{(o)}$, состоящее из так называемых «шаблонных» изображений объекта o. Все элементы $F^{(o)}$

априорно доступны для анализа и представляют собой изображения объекта o, полученные в разных условиях.

Любое изображение $\mathbf{f} \in \mathbf{F}^{(o)}$, $\mathbf{f} = \{\varphi(x)|x \in C[\mathbf{f}] \subseteq R^m\}$, целевого объекта представляет собой совокупность наборов информативных, числовых параметров $\varphi(x)$, вычисленных в точках с координатами $x \in C[\mathbf{f}] \subseteq R^m$. Параметры $\varphi(x)$ изображения $\mathbf{f}$ определены в совокупности точек x, совместно образующих связную, замкнутую область пространства $C[\mathbf{f}] \subseteq R^m$.

$us(x)$, $x \in R^m$ — значение набора информативных, числовых параметров в точке $x \in R^m$, соответствующей местоположению фона или подстилающей поверхности.

Область $C[\mathbf{f}]$ называется **контуром** объектового изображения $\mathbf{f}$. Геометрический размер контура $C[\cdot]$ или его диаметр: $|C[\cdot]| = \operatorname*{Sup}_{x_1,x_2 \in C[\cdot]} \|x_1 - x_2\|$. Можно сказать, что объектовое изображение $\mathbf{f}$ определено только внутри соответствующего ему контура $C[\mathbf{f}]$.

Центр произвольной связной области $U = \{x|x \in U\} \subseteq R^m$ — точка $x_0 \in U$ с декартовыми координатами: $< x_0 >_j^D = (x_{\max,j} - x_{\min,j})/2, 1 \le j \le m$, где $x_{\min,j} = \operatorname*{Min}_{x \in U} <x>_j^D$, $x_{\max,j} = \operatorname*{Max}_{x \in U} <x>_j^D$.

Обозначим: $Cntr[U]$ — операцию определения центра области U: $x_0 = Cntr[U]$.

Соответственно, центр изображения $\mathbf{f} = \{\varphi(x)|x \in C[\mathbf{f}]$ — точка $x_0 = Cntr[C[\mathbf{f}]]$.

Когда центром изображения $\mathbf{f} \in \mathbf{F}^{(o)}$ является точка $x_0 = C[\mathbf{f}(x_0)]$, это записывается так: $\mathbf{f}(x_0) = \{\varphi(x|x_0)|x \in C[\mathbf{f}(x_0)]\}$.

Значение изображения $\mathbf{f}(x_0)$ в точке $x \in C[\mathbf{f}(x_0)]$: $\mathbf{f}(x_0)(x)$.

$[\varphi(x|x_0), r(x|x_0)]$, $\varphi(x|x_0) \in (0,2\pi)$ — полярные координаты произвольной точки $x \in C[\mathbf{f}(x_0)]$ при условии, что начало координат — точка $x_0 \in R^m$.

Для произвольной $x \in C[\mathbf{f}(x_0)]$ обозначим полярные координаты этой точки:

$$\langle x \rangle_\varphi^P = \varphi(x|x_0),\ \langle x \rangle_r^P = r(x|x_0),$$

где $\varphi(x|x_0)$ — угол; $r(x|x_0)$ — радиус точки $x \in C[\mathbf{f}(x_0)]$ в полярной системе координат; P — индекс, обозначающий полярность координатной системы.

$\Xi(x)$ — прямоугольный параллелепипед (Bounding Box) с размерами $|\Xi(x)| = (x_\Xi, y_\Xi)$, для заданных $0 < x_\Xi, y_\Xi < \infty$, с центром в точке $x \in R^m$, $x = Cntr[\Xi(x)]$. Так как $m = 2$, то в дальнейшем будем называть эти параллелепипеды рамками или **поисковыми рамками**. Если рамка $\Xi(x)$ построена на изображении ω, запишем как: $\Xi(x|\omega)$. Если место привязки рамки в конкретном контексте, или изображение, на котором размещена эта рамка, не имеют значения, то вместо $\Xi(x|\omega)$ запишем $\Xi()$ или $\Xi(\cdot|\omega)$.

$\Omega = \{\omega(x)|x \in C[\Omega] \subseteq R^2\}$ — изображение, в общем случае представляющее из себя сцену, в том числе содержащую и искомое изображение $\mathbf{f}_l(x_0)$. Таким образом, кроме искомого изображения $\mathbf{f}_l(x_0)$, сцена Ω может содержать изображения $\mathbf{f}_k(x) = \{\varphi_k(x)|\cdot\} \notin \mathbf{F}^{(o)}$ иных, нецелевых объектов или фон $\{us(x)\}$.

Границы контура $C[\mathbf{f}(x_0)]$: $\beta[\mathbf{f}(x_0)] = \{[\varphi, r_\varphi(x_0)]|\varphi \in (0,2)\}$, где $r_\varphi(x_0) = \operatorname*{Max}_{x \in C[\mathbf{f}(x_0)]} \langle x \rangle_r^P$ для произвольного $\varphi \in (0,2\pi)$.

$\mathbf{f}_l(x_0) = \{\varphi_l(x|x_0)|x \in C[\mathbf{f}_l(x_0)]\} \in \mathbf{F}^{(o)} \backslash F^{(o)}$ — искомое изображение объекта o, такое, что $C[\mathbf{f}_l(x_0)] \subset C[\Omega]$: $C[\mathbf{f}_l(x_0)] \cap C[\Omega] = C[\mathbf{f}_l(x_0)]$.

На сцене Ω задана декартова система координат (координатная плоскость), в которой производятся все координатные преобразования и расчеты.

$\mathbf{\Xi}$ — множество всех возможных поисковых рамок. Для простоты, но без потери общности, считаем, что стороны любой рамки $\Xi \in \mathbf{\Xi}$ параллельны осям координатной плоскости сцены Ω. В общем случае возможны варианты произвольной ориентации рамок по углу поворота. Данный случай будет рассмотрен в разделе «Исследование алгоритма на реальных изображениях».

Координаты углов рамки $\Xi(x)$ составляют множество пар $A^{[\Xi(x)]} = \{a_i^{[\Xi(x)]}\} \subseteq R^2$.

Когда говорится, что некое изображение ω помещено в рамку Ξ, это означает то, что $C[\omega] \subseteq \Xi$, и записывается следующим образом: $\Xi\{C[\omega]\}$.

В общем случае рамка $\Xi(x|\omega)$ может служить для выделения некоторого фрагмента сцены ω в окрестности точки x. В этом случае рамка может не содержать полных контуров изображений $\mathbf{f} \in \mathbf{F}^{(o)}$, напротив — включать в себя лишь фрагменты контуров изображений $\mathbf{f}$ или фона $us(\cdot)$. Рамки такого типа для удобства условимся называть **фоновыми**, так как они служат для ограничения области фоновых изображений.

$A(\theta, \Delta, s)$ — оператор линейного конформного преобразования изображения, иначе называемый «angle-preserving conformal map»: $R^2 \xrightarrow{A(\theta,\Delta,s)} R^2$, который преобразует изображение $f \in \mathbf{F}^{(o)}$ в $f' \in \mathbf{F}^{(o)}$: $f' = A(\theta, \Delta, s)f$ с сохранением углов. Здесь параметры: θ — угол поворота, Δ — сдвиг и s — масштаб.

Рамка называется «идеальной фиксирующей» для $\mathbf{f}(x_0)$ и обозначается $\Xi_\mathbf{f}$, если следующее высказывание верно:

$$\exists \Xi' \in \mathbf{\Xi}: |\Xi'\{C[\mathbf{f}(x_0)]\} \backslash \{C[\mathbf{f}(x_0)]| < \\ < |\Xi_\mathbf{f}\{C[\mathbf{f}(x_0)]\} \backslash C[\mathbf{f}(x_0)]|.$$

Легко показать, что

$$(x_0 = Cntr[C[\mathbf{f}]]) \Leftrightarrow (x_0 = Cntr[\Xi_\mathbf{f}(x|\cdot)\{C[\mathbf{f}]\}]),$$

$$\forall \Xi_\mathbf{f} \in \Xi \exists \mathbf{f}' \ne \mathbf{f} \in \mathbf{F} : \Xi_\mathbf{f} = \Xi_{\mathbf{f}'} \text{ и } \forall \mathbf{f} \in \mathbf{F} : \exists! \Xi_\mathbf{f} \in \mathbf{\Xi}.$$

Рамка называется «неидеальной фиксирующей» для $\mathbf{f}(x_0)$ и обозначается символом $\Xi_{\neg\mathbf{f}}$, если $\Xi \in \mathbf{\Xi}: (\Xi \ne \Xi_\mathbf{f}) \wedge (Cntr[C[\mathbf{f}]] = Cntr[\Xi\{C[\mathbf{f}]\}] = x_0)$. На рис. 1 рамка с маркером *2* является «идеальной» для изображения автомобиля, а рамки с маркерами *1* и *3* — примеры «неидеальных» рамок для целевого изображения. Маркером x_0 отмечен искомый центр изображения автомобиля. Для случая произвольной ориентации рамок «идеальная» рамка будет иметь угол поворота, совпадающий с углом разворота изображения. Стоит обратить внимание на нечеткость, слабую контрастность и размытость контуров изображения. Но именно поиск и оценка центра изображений такого качества является целью настоящей работы.

Опорные изображения объекта o — изображения $\mathbf{f}_{sp}^{(l)} \in F^{(o)}$, $l \in \mathbf{L} = \{1, \ldots, L\}$, $|\mathbf{L}| < \infty$, такие как $|\Xi_{\mathbf{f}_{sp}^{(l)}}| = (x_{sp}, y_{sp})$, где размеры (x_{sp}, y_{sp}), $0 < x_{sp}, y_{sp} < \infty$ определены априорно.

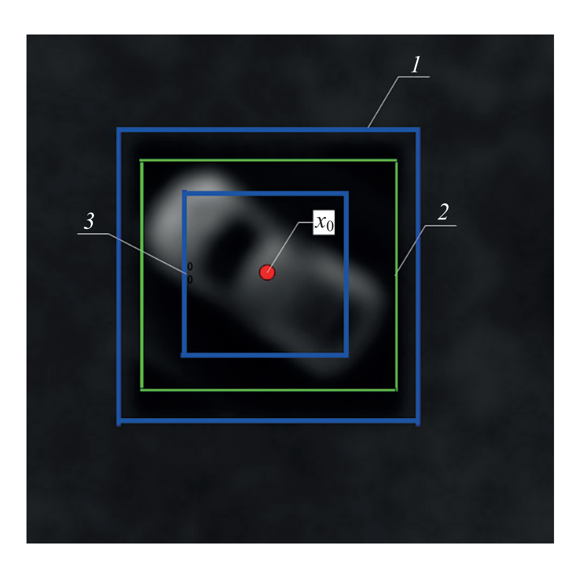


Рис. 1. Примеры рамок

Fig. 1. Examples of frames

Масштаб $s_{\mathbf{f}}$ произвольного изображения $\mathbf{f} \in F^{(o)}$ — кортеж $s_{\mathbf{f}} = (s_{\mathbf{f}}^{(x)}, s_{\mathbf{f}}^{(y)}) = (x_{\Xi_{\mathbf{f}}}/x_{sp}, y_{\Xi_{\mathbf{f}}}/y_{sp})$; $0 < s_{\mathbf{f}}^{(x)}, s_{\mathbf{f}}^{(y)} < \infty$.

$$\forall \mathbf{f} \in F^{(o)} \exists \theta_{\mathbf{f}}, \Delta_{\mathbf{f}}, s_{\mathbf{f}}\text{: } \mathrm{f} = A(\theta_{\mathbf{f}}, \Delta_{\mathbf{f}}, s_{\mathbf{f}})\mathbf{f}_{sp},$$

где $\theta_{\mathbf{f}}$, $\Delta_{\mathbf{f}}$, $s_{\mathbf{f}}$ — угол поворота, сдвиг и масштаб изображения $\mathbf{f}$ относительно $\mathbf{f}_{sp}$ соответственно.

Контур $C[\omega]$, ограничивающий фрагмент фонового изображения ω, совпадает с его рамкой: $C[\omega] = \Xi(\cdot|\omega)$, $x_\omega = Cntr[C[\omega]] = Cntr\,[\Xi(\cdot|\omega)]$. Для объектовых изображений $\mathbf{f} \in \mathbf{F}^{(o)}$, в общем случае $C[\mathbf{f}] \neq \Xi_{\mathbf{f}}$.

Таким образом, любая рамка может быть либо объектовой, либо фоновой, в зависимости от того, контур какого изображения она ограничивает.

Масштаб идеальной рамки $\Xi_{\mathbf{f}}$ — параметр $s = s_{\mathbf{f}}$.

Множество $\boldsymbol{\Xi}$ имеет следующее ограничение:

$$\forall \mathbf{f} \in F^{(o)} \exists \Xi \in \boldsymbol{\Xi} : \Xi = \Xi_{\mathbf{f}}.$$

Сцена $\Omega = \{\omega(x)|x \in C[\Omega]\}$ наблюдается с аддитивным случайным шумом $\{\xi(x)\}$. Если для некоторого $\mathbf{f}_i \in \mathbf{F}^{(o)}$ и $C[\mathbf{f}_i] \subseteq C[\Omega]$, то $\forall x \in C[\mathbf{f}_i]$: $\omega(x) = \varphi_i(x) + \xi(x)$.

Зашумленный фрагмент сцены $\Omega' = \{\varphi(x|x_0) + \xi(x)|x \in C[\Omega] \cap C[\mathbf{f}(x_0)]\} \subseteq \Omega$, где $\mathbf{f}(x_0) = \{\varphi(x|x_0)\}$, $\mathbf{f}(x_0) \in \mathbf{F}^{(o)}$, является $\boldsymbol{\varepsilon}$-гладким, если для некого $\boldsymbol{\varepsilon} > 0$ имеем:

$$\forall (x \in \beta[\mathbf{f}(x_0)], \Delta \neq 0) : |\mathbf{E}(\varphi(x|x_0) + \xi(x) - \varphi(x + \Delta|x_0) - \xi(x + \Delta))| \leq \mathbf{E}\xi^2(x)(1 + \boldsymbol{\varepsilon}||\Delta||).$$

Здесь и далее $\mathbf{E}$ — оператор математического ожидания.

Определение. *Оценка x^+ величины x_0 (центр изображения $\mathbf{f} \in \mathbf{F}^{(o)}$) называется R-робастной по отношению к шумовой составляющей $\xi(x)$, если для любых двух ситуаций, первая из которых (оценка x_1^+) соответствует величине $\mathbf{E}\xi_1^2(x)$, а вторая (оценка x_2^+) величине $\mathbf{E}\xi_2^2(x)$, и $\exists R$:$(1 < R < \infty$, $\mathbf{E}\xi_2^2(x)$, $\mathbf{E}\xi_1^2(x) \leq R$, имеет место следующая импликация*:

$$(\mathbf{E}\xi_2^2(x)/\mathbf{E}\xi_1^2(x) > 1) \Rightarrow$$
$$\Rightarrow (\mathbf{E}\xi_2^2(x)/\mathbf{E}\xi_1^2(x) \geq \mathbf{E}||x_2^+ - x_0||/\mathbf{E}||x_1^+ - x_0||). \quad (1)$$

Постановка задачи

Дано:

1. Множество $F^{(o)} = \cup_s f_s^{(o)}$:$(\mathbf{f}_i \in f_s^{(o)}$:$\mathbf{f}_i = \{\varphi_i(x)|x \in C[\mathbf{f}_i]\})$.
2. Сцена Ω, возможно содержащая целевое изображение $\mathbf{f}_l(x_0)$.

   Для неизвестных целевого изображения $\mathbf{f}_l(x_0)$ и параметров линейного конформного преобразования θ_l, Δ_l, s_l верна следующая запись:

$$\exists(\theta_l \in (0{,}2\pi); \Delta_l, s_l \in R^2, l \in \mathbf{L}) : \mathbf{f}_l(x_0) = A(\theta_l, \Delta_l, s_l)\mathbf{f}_{sp}^{(l)}(x),$$
$$\forall x \in C[\Omega] \cap C[\mathbf{f}_l(x_0)] : \omega(x) = \varphi_l(x|x_0) + \xi(x),$$
$$\forall x \in C[\Omega] \backslash C[\mathbf{f}_l(x_0)] \exists\, \mathbf{f}_k(x) = \{\varphi_k(x)|\cdot\} \notin \mathbf{F}^{(o)} : (\omega(x) =$$
$$= \xi(x)) \vee (\omega(x) = \varphi_k(x) + \xi(x)),$$

где $\xi(x)$ — аддитивный пространственно-коррелированный стохастический шум наблюдения с неизвестным распределением, матрица корреляции $\mathbf{E}\xi^2(x) < \infty$ которого неизвестна.

3. В общем случае $\mathbf{f}_l(x_0) \notin F^{(o)}$.
4. $\forall x \in C[\Omega] \cap C[\mathbf{f}_l(x_0)] : \omega(x) = \varphi_l(x|x_0) + \xi(x)$.

Необходимо:

1) построить процедуру обнаружения $Search(\Omega)$ целевого изображения $\mathbf{f}_l(x_0)$ на сцене Ω, а также точечную оценку x^+ его центра x_0; R-робастную (R — некоторая конечная, положительная константа) к аддитивным, коррелированным шумам $\xi(x)$ при условии априорной неопределенности относительно параметров линейного конформного преобразования θ_l, Δ_l, s_l:

$$Search(\Omega) = \begin{cases} x^+, \ if \ \ C[\mathbf{f}_l(x_0)] \cap \Omega = C[\mathbf{f}_l(x_0)]. \\ 0, \ \ if \ \ C[\mathbf{f}_l(x_0)] \cap \Omega = \varnothing \end{cases}.$$

2) для заданной величины $0 < \alpha < 1$ определить способ построения конечного множества Υ, $||\Upsilon|| < \infty$, такого что

$$\mathbf{P}(x_0 \in \Upsilon(x^+)) \geq 1 - \alpha.$$

Метод решения

Для обеспечения стабильности работы алгоритма в условиях коррелированных шумов, нечеткости целевого изображения и неоднородностей сцены, а также для обеспечения устойчивости к влиянию линейных конформных преобразований (сдвиг, поворот и масштаб), решение поставленной задачи сводится к решению задачи классификации (возможно мультиклассовой) в экстремальной постановке, при условии специального способа формирования как обучающих множеств, так и самой поисковой процедуры. Как отмечалось в разделе «Идея метода», бинарная классификация изображения конкретного участка сцены эквивалентна задаче обнаружения (детекции) искомого изображения на сцене, а процедура поиска координат точки максимума функции правдоподобия бинарного классификатора формирует оценку координат центра искомого изображения.

Рассмотрим два множества.

1. Множество $\mathbf{\Omega}^{(0)}$ (нулевой класс):

   $\mathbf{\Omega}^{(0)} = \{\Omega_i^{(0)}(x_0)\}$ составляют различные изображения $\mathbf{f}_i = \{\varphi_i(x|x_0)\} \in F^{(o)} \backslash \mathbf{f}_l(x_0)$ целевого объекта o, на фоне

$us(x)$, при наличии пространственных, коррелированных шумов $\xi(x)$, помещенные в соответствующие идеальные рамки $\Xi_{\mathbf{f}}(x_0|\cdot)$ с центрами в точке x_0.

$$\forall\Omega_i^{(0)}(x_0) \in \mathbf{\Omega}^{(0)}: \Omega_i^{(0)}(x_0) = \{\omega_i(x)\},$$

$$\omega_i(x) = \begin{cases} \varphi_i(x|x_0) + \xi(x),\ if\ x \in C[\mathbf{f}_i(x_0)] \\ us(x) + \xi(x),\ if\ x \in \Xi_{\mathbf{f}}\{C[\mathbf{f}_i(x_0)]\}\backslash C[\mathbf{f}_i(x_0)] \end{cases},$$

$$Cntr[\Xi_{\mathbf{f}}\{C[\mathbf{f}_i]\}] = x_0.$$

Если изображение $\Omega_i^{(0)}(x_0) \in \mathbf{\Omega}^{(0)}$ в качестве своего фрагмента содержит $\mathbf{f}_i(x_0)$, будем говорить, что изображение $\Omega_i^{(0)}(x_0)$ основано на $\mathbf{f}_i(x_0)$ и, для удобства символьной записи, обозначать это следующим образом:

$$\Omega_i^{(0)}(x_0) = \mathbf{B}[\mathbf{f}_i(x_0)],$$

где $\mathbf{B}$ — обозначение того, что изображение $\Omega_i^{(0)}(x_0)$ основано на $\mathbf{f}_i(x_0)$. Изображение $\Omega_i^{(0)}(x_0) \in \mathbf{\Omega}^{(0)}$ может быть основано только на одном изображении $\mathbf{f}_i(x_0)$.

Координаты точек изображений $\Omega_i^{(0)} \in \mathbf{\Omega}^{(0)}$ определены в одной и той же координатной плоскости, которую назовем единой.

$\mathbf{\Omega}^{(0)}$ может быть разбито на «равномасштабные» подмножества: $\{W_s^{(0)}\}$: $(\cup_s W_s^{(0)} = \mathbf{\Omega}^{(0)}, \cap_s W_s^{(0)} = \varnothing)$, $\forall_s\forall\mathbf{B}[\mathbf{f}_1(x_0)], \mathbf{B}[\mathbf{f}_2(x_0)] \in W_s^{(0)}$: $(\theta_{\mathbf{f}_1} \neq \theta_{\mathbf{f}_2}) \vee (s_{\mathbf{f}_1} = s_{\mathbf{f}_2} = s)$.

2. Множество $\mathbf{\Omega}^{(1)}$ (первый класс):

$\mathbf{\Omega}^{(1)} = \{\Omega_i^{(1)}\}$ составляют связные множества $c_i \subset C[\mathbf{f}_i(x_0)]$, $\mathbf{f}_i = \{\varphi_i(x|x_0)\} \in F^{(o)}$, на фоне $us(x)$, при наличии пространственных, коррелированных шумов $\xi(x)$, помещенные в соответствующие «фоновые» рамки $\Xi(x_i)$ различного размера (масштаба).

Множество $c_i \subset \omega_i$ таково, что c_i: $(\varepsilon_1 < |C[\mathbf{f}_i(x_0)]\backslash c_i| \leq \leq \varepsilon_2,\ 0 < \varepsilon_1,\ \varepsilon_2 < \infty) \vee (x_0 \notin c_i)$.

$$\forall(\Omega_i^{(1)} \in \Omega^{(1)}, \Omega_i^{(1)} = \{\omega_i(x)\})\exists\Xi(x_i): \left(\forall x \in \Xi(x_i): \omega_i(x) = = \begin{cases} \varphi_i(x|x_0) + \xi(x),\ if\ x \in c_i \\ us(x) + \xi(x),\ if\ x \in \Xi(x_i)\backslash c_i \end{cases}\right).$$

По способу построения $\mathbf{\Omega}^{(1)}$ состоит из участков изображений $\mathbf{f} \in F^{(o)}$, помещенных в «фоновые» рамки, образующих множество $\mathbf{C} = \{\Xi(x_i)|\mathbf{f})\}$ таким образом, чтобы ни одна рамка из $\mathbf{C}$ не захватывала центр изображения x_0.

Способ формирования множества фоновых рамок $\mathbf{C}$ — критически важен, поэтому рассмотрим его подробнее. Рассмотрим окружность $Cir(x_0, r)$ с центром в точке $x_0 \in R^m$ и радиусом r, при $r > 0$, а также последовательность положительных величин: $\Phi = \{\varphi_i|i = 1, \ldots, n\}$, $\sum_{i=1}^{n}\varphi_i = 2\pi$, $\underset{1\leq i\leq n-1}{\forall}||\varphi_i - \varphi_{i+1}|| = 2\pi/n$. Пусть $[\varphi_i, r]$, $\varphi_i \in \Phi$ — полярные координаты некоторой точки $[\varphi_i, r] \in Cir(x_0, r)$, $D[\varphi_i, r]$ — координаты этой точки в единой координатной плоскости. Множество $\mathbf{C}$ рассмотрим в виде множества $\mathbf{C}_{sz}^{(r,\Phi)} = \{\Xi_{sz}^{\varphi_i}|\varphi_i \in \Phi\}$, представляющем собой множество рамок равного размера $sz(\underset{i}{\forall} sz = |\Xi_{sz}^{\varphi_i}|)$ такое, что

$$\mathbf{C}_{sz}^{(r,\Phi)}: \forall\varphi_i \in \Phi\exists!\Xi = \Xi_{sz}^{\varphi_i} \in \mathbf{C}_{sz}^{(r,\Phi)}: (\exists a \in A^{[\Xi]}, [\varphi_i, r] \in \in Cir(x_0, r): a = D[\varphi_i, r]).$$

Очевидно, что $\forall\Xi_{sz}^{\varphi_i} \in \mathbf{C}_{sz}^{(r,\Phi)}:|\Xi_{sz}^{\varphi_i} \cap Cir(x_0, r)| = 1$.

Множество $\mathbf{\Omega}^{(1)}$, подобно множеству $\mathbf{\Omega}^{(0)}$, разбито на «равномасштабные» подмножества $\{W_s^{(1)}\}$, такие как $(\cup_s W_s^{(1)} = \mathbf{\Omega}^{(1)}, \cap_s W_s^{(1)} = \varnothing)$ и

$$\underset{s}{\forall}(\forall c_1 \subset \Omega_1^{(1)} = \mathbf{B}[\mathbf{f}_1] \in W_s^{(1)}, c_2 \subset \Omega_2^{(1)} = \mathbf{B}[\mathbf{f}_2] \in W_s^{(1)}): :(\theta_{\mathbf{f}_1} \neq \theta_{\mathbf{f}_2}) \vee (s_{\mathbf{f}_1} = s_{\mathbf{f}_2} = s).$$

Множество $W_s^{(0)} \cup W_s^{(1)}$ назовем **уровнем масштаба *s***.

Пусть $\chi[\Xi(x|\omega)] \in \mathbf{X} \subseteq R^p$ — некая p-мерная функция, вычисленная по изображению ω, помещенному в рамку $\Xi(x|\omega)$. Будем называть эту функцию вектором характеристических параметров (вектором признаков) изображения в рамке $\Xi(x|\omega)$. Как упоминалось ранее, рамка может быть либо фоновой, либо объектовой. По значению $\chi[\Xi(x|\omega)]$ можно определять тип рамки, т. е. ответить на вопрос: ограничивает ли эта рамка целевой объект, либо она ограничивает фон? Собственно, эта простая идея и лежит в основе предлагаемого в данной работе подхода.

Теперь рассмотрим множество бинарных классификаторов $\{CL_s\}$: $\forall s$: CL_s: $\mathbf{X} \xrightarrow{CL_s} \{0,1\}$. Каждый CL_s настроен для работы с рамкой Ξ: $\exists\mathbf{f} \in F^{(o)}$: Ξ: $\Xi_{\mathbf{f}}$, которая является идеальной для некоторого $\mathbf{f} \in F^{(o)}$, и поэтому имеет масштаб $s = s_{\mathbf{f}}$. В этом случае каждый CL_s обучается на множествах $\{(\chi[\Xi(x_0|w_{0,s})], 0)| w_{0,s} \in W_s^{(0)}\}$ — нулевой класс и $\{(\chi[\Xi(x|w_{1,s})], 1)| w_{1,s} \in W_s^{(1)}\}$ — первый класс, т. е. классификатор CL_s обучен на уровне масштаба s. Иначе: $CL_s(\chi[\Xi(x|w)]) = k$, $k \in \{0,1\}$, k — номер класса: это решение, которое выносит классификатор $CL_z(\chi[\Xi(x|w)])$ по значению $\chi[\Xi(x|w)]$.

$$\underset{s}{\forall}\exists\rho_{CL_s}(\chi) \in (0,1), \chi \in \mathbf{X}: \forall\chi[\Xi(x|\cdot)] \in \mathbf{X}: \rho_{CL_s}(\chi[\Xi(x|\cdot)]) \sim \sim \mathbf{P}(C[\mathbf{f}(x_0)] \subset \Xi(x|\cdot)). \quad (2)$$

Здесь для $\chi[\Xi(x|\mathbf{f})] \in \mathbf{X}$ величина $\mathbf{P}(C[\mathbf{f}(x_0)] \subset \Xi(x|\mathbf{f}))$ представляет собой байесовскую вероятность того, что контур $\Xi(x|\mathbf{f})$, с центром в точке x, содержит контур изображения целевого объекта $\mathbf{f}$ и центр контура $x = x_0$. Причем:

$$(x \neq x_0) \Rightarrow (\mathbf{P}(C[\mathbf{f}(x_0)] \subset \Xi(x_0|\mathbf{f})) > > \mathbf{P}(C[\mathbf{f}(x_0)] \subset \Xi(x|\mathbf{f}))). \quad (3)$$

Из импликации (3) сразу следует, что

$$x_0 = \mathrm{Arg}\sup_{x\in C[\Omega^*]} \mathbf{P}(C[\mathbf{f}(x_0)] \subset \Xi(x|\mathbf{f})). \quad (4)$$

Заменим в (4) величину $\mathbf{P}(C[\mathbf{f}(x_0)] \subset \Xi(x|\mathbf{f}))$ на ее приближение $\rho_{CL_s}(\chi[\Xi(x|\mathbf{f})])$, например, на широко известные масштабирование Платта или softmax scaling [10], и рассмотрим оценку x_0^* в следующем виде:

$$x_0^* = \mathrm{Arg}\sup_{x\in C[\Omega^*]} \rho_{CL_s}(\chi[\Xi(x|\mathbf{f})]).$$

В дальнейшем для простоты вместо $\rho_{CL_s}(\chi[\Xi(x|\mathbf{f})])$ и $CL_s(\chi[\Xi(x|\mathbf{f})])$ условимся использовать более компактные обозначения: $\rho_{CL_s}(x)$ и $CL_s(x)$ соответственно.

С учетом (2) имеем $x_0^* \sim x_0$. Таким образом, x_0^* — оценка искомой величины x_0, а величина $\rho_{CL_s}(x_0^*)$ ха-

рактеризует качество этой оценки, так как исходя из выражения (2):

$$\rho_{CL_s}(x_0^*) \sim \mathbf{P}(C[\mathbf{f}(x_0)] \subset \Xi(x|\mathbf{f})).$$

На основании способа обучения CL_s не может быть эффективен для тех изображений $\mathbf{f}'$, у которых $s_{\mathbf{f}'} \neq s$. Для того, чтобы преодолеть эту очевидную особенность в условиях произвольного масштаба классифицируемых изображений, используется не один классификатор, а **комбинация разномасштабных классификаторов**, представляющая собой взвешенную сумму следующего вида:

$$CL_s(x) = \sum_{s \in S} CL_s(x)\gamma(\rho_{CL_s}(x)), \tag{5}$$

где $\sum_{s \in S} \gamma(\rho_{CL_s}(x)) = 1$, $\gamma(\cdot) > 0$, $\gamma(\cdot)$ — строго монотонная, возрастающая функция; S — дискретное множество допустимых масштабов.

В этом случае для оценки $\mathbf{P}(C[\mathbf{f}(x_0)] \subset \Xi(x_0|\mathbf{f}))$ используется не $\rho_{CL_s}(x)$, а величина

$$\boldsymbol{\rho}_s(x_0^{(S)}) = \sum_{s \in S} \rho_{CL_s}(x_0^{(S)})\gamma(\rho_{CL_s}(x_0^{(S)})),$$

где

$$x_0^{(S)} = \text{Arg} \sup_{x \in C[\Omega^*]} \boldsymbol{\rho}_s(x).$$

Классификационное решение при использовании (5) принимается по правилу: $[CL_s]$, т. е. по правилу округления к ближайшему целому. В качестве функции $\gamma(\cdot)$ можно использовать $\gamma(\rho_{CL_s}(x)) = \rho_{CL_s}(x)(\sum_{k \in S} \rho_{CL_k}(x))^{-1}$.

Выбор множества S, которое определяет априорно просчитанные варианты значений масштаба, описан в разделе «Стабилизация алгоритма…».

Рассмотрим функционал:

$$\boldsymbol{\Delta}(x|\tau) = 1 - \boldsymbol{\rho}_S(x) + \tau CL_s(x),\ \tau > 0. \tag{6}$$

Очевидно, что $\boldsymbol{\Delta}(x_0^{(S)}|\tau) = \inf_x \boldsymbol{\Delta}(x|\tau)$ при условии $CL_s(x_0^{(S)}) = 0$. При корректном обучении классификатора CL_s равенство $CL_s(x_0^{(S)}) = 0$ будет достигаться с высокой вероятностью. В выражении (6) слагаемое $\tau CL_s(x)$ выполняет роль штрафной функции. Штрафуются те значения переменной x, при которых классификатор указывает на принадлежность изображения внутри рамки Ξ к классу $\Omega^{(1)}$ (фон или фрагменты нецелевых изображений). Величина τ выполняет роль регуляризирующей константы и подбирается в процессе настройки.

Окончательно точечной оценкой величины x_0 будем считать величину

$$x_0^+ = \text{Arg} \inf_{x \in C[\Omega^*]} \boldsymbol{\Delta}(x|\tau). \tag{7}$$

Очевидно, что

$$\forall(\mathbf{f} \in \mathbf{F}^{(o)}, x \in C[\mathbf{f}(x_0)]): \boldsymbol{\rho}_s(x) = \mathbf{E}(\boldsymbol{\rho}_s(x)) + \zeta(x),$$

где $\zeta(x)$ — случайный компонент; $\mathbf{E}\zeta(x) = 0$.

Для заданного $0 < \alpha < 1$ рассмотрим двусторонний квантиль $-c_\alpha, c_\alpha$, $\mathbf{P}(-c_\alpha \leq \zeta(x) \leq c_\alpha) \geq 1 - \alpha$, $(\mathbf{P}(|\zeta(x)| \leq c_\alpha) \geq \geq 1 - \alpha)$, который вычисляется на стадии предварительного численного анализа, также как и величины $\mathbf{E}(\boldsymbol{\rho}_s(x_0)) = \boldsymbol{\delta} > 0$ и c_α. Значение c_α зависит от свойств изображения, от способа вычисления величины $\chi[\Xi(x|\cdot)$, а также от статистических характеристик шумов $\xi(x)$.

Теорема. *Рассмотрим множество:*

$$\Upsilon = \{x|\boldsymbol{\rho}_S(x) - \boldsymbol{\delta} \leq c_\alpha.$$

Пусть $\mathbf{E}(\boldsymbol{\rho}_S(x_0)) = \max_x(\mathbf{E}(\boldsymbol{\rho}_S(x))) = \boldsymbol{\delta}$.

Тогда $\mathbf{P}(x_0 \in \Upsilon) \geq 1 - \alpha$.

Доказательство. Из условия теоремы имеем:

$$(x = x_0) \Leftrightarrow (\mathbf{E}(\boldsymbol{\rho}_S(x)) = \boldsymbol{\delta}). \tag{8}$$

Обозначим: $\varepsilon_\zeta = c_\alpha - \zeta(x)$. Очевидно, что

$$\forall x \in \Upsilon: (|\zeta(x)| \leq c_\alpha) \Rightarrow (\varepsilon_\zeta \geq 0). \tag{9}$$

Далее:

$$\forall x \in \Upsilon: (|\zeta(x)| \leq c_\alpha) \Rightarrow (\boldsymbol{\rho}_S(x) - \boldsymbol{\delta}) =$$
$$= \mathbf{E}(\boldsymbol{\rho}_S(x)) - \boldsymbol{\delta} + \zeta(x) \leq c_\alpha) \Rightarrow \mathbf{E}(\boldsymbol{\rho}_S(x)) - \boldsymbol{\delta} \leq \varepsilon_\zeta. \tag{10}$$

При $(|\zeta(x)| \leq c_\alpha)$, $x \in \Upsilon$, из условия теоремы следует, что значения величины $(\mathbf{E}(\boldsymbol{\rho}_S(x)) - \boldsymbol{\delta})$ могут принимать любое значение из интервала $E_\zeta = (-\infty, \varepsilon_\zeta)$, причем, с учетом (9), $0 \in E_\zeta$, поэтому (10) можно записать в следующем виде:

$$\forall x \in \Upsilon: (|\zeta(x)| \leq c_\alpha) \Rightarrow (\mathbf{E}(\boldsymbol{\rho}_S(x)) - \boldsymbol{\delta}) \in E_\zeta) \Rightarrow$$
$$\Rightarrow (\mathbf{E}(\boldsymbol{\rho}_S(x)) - \boldsymbol{\delta} = 0 \in \varepsilon_\zeta). \tag{11}$$

Из (8) и (11) имеем:

$$\forall x \in \Upsilon: (|\zeta(x)| \leq c_\alpha) \Rightarrow (\mathbf{E}(\boldsymbol{\rho}_S(x)) - \boldsymbol{\delta} = 0 \in E_\zeta) \Rightarrow$$
$$\Rightarrow (x_0 \in \Upsilon). \tag{12}$$

Из (12) следует, что:

$$(\mathbf{P}(|\zeta(x)| \leq c_\alpha) \geq 1 - \alpha) \Rightarrow (\mathbf{P}(x_0 \in \Upsilon) \geq 1 - \alpha).$$

Теорема доказана.

Предложенный метод формирования доверительного множества во многом аналогичен методу, предложенному в [11–13]. Результаты этой теоремы состоят не только в способе определения доверительного множества для величины x_0, но и в способе принятия решения о том, является ли $\Xi(x|\cdot)$ идеальной рамкой для целевого изображения? Иначе говоря: насколько вероятно то, что $\Xi(x|\cdot)$ содержит большую часть контура $C[\mathbf{f}_1(x_0)]$ целевого изображения $\mathbf{f}_I(x_0)$? Правило простое: $(\boldsymbol{\rho}_S(x) < c_\alpha + \boldsymbol{\delta}) \Rightarrow (\mathrm{P}(\mathbf{f}_I(x_0) \not\subseteq \Xi(x|\cdot)) \geq 1 - \alpha)$.

Выбор вектора признаков

Вектор признаков $\chi[\Xi(x|\omega)] \in \mathbf{X} \subseteq R^p$ вычисляется по части изображения ω, которое попало внутрь рамки Ξ. С одной стороны, величина p (размерность χ) должна быть минимизирована с целью снижения требований к мощности вычислителя. С другой стороны, слишком малое значение p понизит надежность обнаружения целевого изображения. Как правило, изображение имеет ключевые точки, которые в значительной мере характеризуют его индивидуальность, а следовательно,

и распознаваемость. Потому для конкретного изображения желательно выделять эти точки и формировать вектор χ в их окрестностях. Например, это могут быть границы контура, углы или иные значимые особенности. Для выделения ключевых точек традиционно используются такие методы как SIFT, FAST, ORB, SURF. Однако многочисленные эксперименты показали, что все эти подходы теряют работоспособность в условиях интенсивных коррелированных помех и размытых контуров. В настоящей работе рассмотрены именно такие условия. В качестве альтернативы можно найти информативные области изображения с использованием кластерного анализа либо метода главных компонент. Но эти методы довольно затратны в вычислительном аспекте, что вступает в противоречие с необходимостью обеспечения функционирования системы в реальном масштабе времени. В рамках настоящей работы χ представлен в следующем виде:

$$\chi[\Xi(x|\omega)] = (\Xi(x|\omega)[x_1], \ldots, \Xi(x|\omega)[x_p]) \in R^p,$$

$$\forall_{1 \le i \le p}\ x_i \in \Xi(x|\omega): \|x_{i+1} - x_i\| = \theta,$$

где $\theta > 0$, $\Xi(x|\omega)[x_i] = \omega(x_i)$; $p \le x_\Xi y_\Xi$; $|\Xi(x|\omega)| = (x_\Xi, y_\Xi)$.

С одной стороны, при малом p, вычисление χ указанным образом предъявляет минимальные требования к мощности вычислителя, необходимой для его определения в реальном масштабе времени. Но с другой, — большие значения размерности χ, приближающиеся к $x_\Xi y_\Xi$, обуславливают повышение вычислительной нагрузки при решении задачи (7). Исходя из этого, выбор параметра p (количество тестовых точек) представляет собой крайне важную задачу и осуществляется на этапе настройки системы на множество целевых изображений. В значительной степени на величины параметра p оказывает значение радиуса корреляции изображения, который определяет ширину автокорреляционной функции вблизи ее максимума и характеризует гладкость изображения. Следует заметить, что чем более гладким является изображение, тем меньше локальных экстремумов у функционала $\boldsymbol{\Delta}(x|\tau)$, и тем проще решение задачи (7). Но, с другой стороны, гладкие изображения — неинформативны, системе поиска целевого изображения не за что «зацепиться», что приводит к снижению надежности поисковой процедуры. Результаты численного исследования влияния значения параметра p на эффективность работы поисковой системы приведены в разделе «Исследование алгоритма на реальных изображениях».

Классификатор

В качестве алгоритма классификации использован классический SVM (Support Vector Machine) [14], который в сравнении с альтернативами (Gradient Boosting [15, 16], XGBoost [17]), обеспечил высокую продуктивность и надежность. Также для решения задачи классификации может быть использована и конволюционная нейронная сеть, но для ее обучения потребуется большая мощность обучающих корпусов, поэтому этот вариант реализации классификатора в настоящей работе не рассмотрен.

Стабилизация алгоритма: неформальное описание способов обеспечения устойчивости к линейным конформным преобразованиям

Сцена, на которой предстоит найти объект **f**, может быть зашумлена пространственно-коррелированной помехой, но, и это значительно усложняет задачу, объект может присутствовать на сцене в неизвестном масштабе и с неизвестным углом поворота. Формально основные методы обеспечения устойчивости метода к неопределенности такого рода были описаны в разделе «Метод решения», а в этом разделе будут приведены некоторые неформальные пояснения их сути.

Обеспечение устойчивости поиска к влиянию пространственно-коррелированной помехи. Поисковая процедура призвана минимизировать (оптимально до нуля) сдвиг между центром x_0 искомого изображения $\mathbf{f}_I(x_0)$ на сцене Ω и центром поисковой рамки. Для обеспечения устойчивости поисковой процедуры совместно использовались два подхода.

Первый подход касается самого способа формирования обучающих множеств $\boldsymbol{\Omega}^{(0)}$ и $\boldsymbol{\Omega}^{(1)}$, используемых при обучении системы поиска. Искомый класс, представленный элементами из $\boldsymbol{\Omega}^{(0)}$, составляют изображения $\Omega_i^{(0)}(x_0) \in \Omega^{(0)}$, $\Omega_i^{(0)}(x_0) = \mathbf{B}[\mathbf{f}_i(x_0)]$, геометрические центры которых помещены в **центры рамок** $\boldsymbol{\Xi}_{\mathbf{f}_i}(x_0|w_{0,s})$, $\mathbf{f}_i \in F^{(o)}$. Из самого способа построения этого множества следует, что максимальная степень правдоподобия $\rho_{CL_x}(\chi[\Xi(x|\cdot)])$ будет у тех x, которые располагаются достаточно близко к искомой величине x_0.

Второй подход касается фонового класса $\boldsymbol{\Omega}^{(1)}$ и построен так, что за пределами области $Cir(x_0, r)$ все участки изображений $\mathbf{f}_i \in F^{(o)}$ в смеси с шумами и фоном, попавшие в рамки $\Xi(x|w_{1,s})$, такие как $|Cir(x_0, r) \cap \Xi(x|w_{1,s})| \le 1$, относятся к фоновому классу (первый класс). Такой способ построения обучающих множеств, при использовании поисковой процедуры оптимизационного типа, позволяет довольно эффективно оценивать x_0 в условиях пространственно-коррелированной помехи, обеспечивая устойчивость поисковой процедуры. Рис. 2 иллюстрирует способ формирования множеств $\boldsymbol{\Omega}^{(0)}$ и $\boldsymbol{\Omega}^{(1)}$. Здесь белая рамка соответствует классу $\boldsymbol{\Omega}^{(0)}$, черные рамки — классу $\boldsymbol{\Omega}^{(1)}$. Черный пунктирный круг обозначает $Cir(x_0, r)$, черная точка — местоположение центра изображения $x_0 \in R^m$, которое подлежит оцениванию.

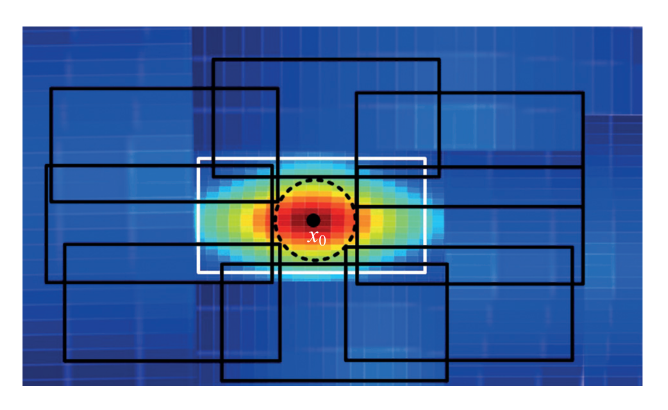


Рис. 2. Формирование обучающих множеств

Fig. 2. Training sets formation

Обеспечение устойчивости к масштабу. Обеспечение устойчивости к масштабу сводится к определению состава множества S, которое назовем масштабной сеткой. Априорно задается диапазон допустимого изменения масштаба s в виде замкнутого интервала $\mathbf{S} = (s_{\min}, s_{\max})$, $s \in \mathbf{S}$. Внутри $\mathbf{S}$ определяется конечное упорядоченное множество точек («сеть»):

$$S = \{s_i | i = 1, \ldots, n_S\}: (s_1 = s_{\min}, s_n = s_{\max},$$
$$\underset{1 \le i \le n-1}{\forall}: (s_{i+1} - s_i) = \upsilon_S = |\mathbf{S}| n_S^{-1}),$$

где υ_S — параметр шага сетки S.

Если $\underset{1 \le i \le n-1}{\forall}: (s_{i+1} - s_i) = \upsilon_S$, то сеть называется регулярной. Численные эксперименты показали, что величина n_S не обязательно должна быть значительной. Например, для практически удовлетворительного функционирования системы обнаружения, при $|\mathbf{S}| = 0{,}25$ (25 % уровень различия в масштабах), достаточно, чтобы $n_S = 5$, т. е. $\upsilon_S = 5\%$. В общем случае для получения практически приемлемых результатов сеть S не обязательно должна быть регулярной. Вопрос оптимального выбора υ_S, представляет собой отдельное исследование, и в значительной степени определяется помеховой обстановкой (матрицей пространственной корреляции шума $\mathbf{E}\xi(x_1)\xi(x_2)$ и его распределением), а также корреляционными свойствами изображений.

Обеспечение устойчивости к повороту. По аналогии с обеспечением устойчивости к масштабу, для устойчивости к повороту вводится множество Θ, которое назовем ротационной сеткой. Множество Θ состоит из априорно определенных значений углов поворота θ_i целевого изображения $\mathbf{f}$ вокруг точки x_0.

$$\Theta = \{\theta_i | i = 1, \ldots, n_\Theta\}: (\alpha_1 = 0, \theta_{n\Theta} = 2\pi,$$
$$\underset{1 \le i \le n-1}{\forall}: (\theta_{i+1} - \theta_i) = \upsilon_\Theta = 2\pi n_\Theta^{-1}),$$

где υ_Θ — шаг сетки Θ.

Для $F^{(o)}$, $\mathbf{L}$ и S допустима следующая запись:

$$\forall(l \in \mathbf{L}, \theta \in \Theta, s \in S)\exists!\mathbf{f}^{(s,\theta)}(x_0) \in F^{(o)}: \mathbf{f}^{(s,\theta)}(x_0) =$$
$$= A(\theta, 0, s)\mathbf{f}_{sp}^{(l)}(x_0),$$

где $\mathbf{L}$ — множество индексов опорных изображений, определенное в разделе «Определения и обозначения».

Мощность ротационной сетки n_Θ принята равной 72, что соответствует шагу углового перебора $\upsilon_\Theta = 5°$. Численные эксперименты показали, что $\upsilon_\Theta = 5°$ оказался достаточным для обеспечения робастности алгоритма к поворотам искомого изображения.

Поисковая процедура

Процедура поиска экстремума (7) является важным пунктом при реализации предлагаемого подхода. Основные требования к поисковой процедуре: устойчивость к шумам, минимизация числа вычислений целевого функционала $\mathbf{\Delta}(x|\tau)$ в процессе поиска, а также высокая вероятность нахождения глобального экстремума. Начальная точка поиска на сцене Ω, в которой предположительно может находиться целевое изображение $\mathbf{f}_l(x_0)$, определена методами грубого обнаружения. Например, для этой цели может использоваться поточечный пороговый классификатор, который принимает решение о том, что на фрагменте присутствует фон в том случае, когда изменение яркости в данной точке изображения во времени меньше некоторого порогового значения. В противном случае принимается решение о том, что в данной части сцены может присутствовать целевой объект. При анализе динамических сцен могут быть использованы различные варианты методов вычитания. Описание этих алгоритмов не рассмотрены в настоящей работы. Так как в общем случае функционал $\mathbf{\Delta}(x|\tau)$ является многоэкстремальным, для обеспечения высокой вероятности схождения к глобальному экстремуму поисковая процедура является двухэтапной.

Задача первого этапа: определение точки x^{Glob} из области притяжения глобального экстремума.

Задача второго этапа: «быстрый» поиск точки глобального экстремума (точки x_0), начиная от точки x^{Glob}.

Для первого этапа функционал $\mathbf{\Delta}(x|\tau)$ модифицируется таким образом, чтобы сделать область притяжения глобального экстремума более «широкой», пусть и с потерей точности. Для этой цели на этапе обучения в состав множества $W_s^{(0)}$ искусственно добавляют образцы $w_{0,s}^{(i)} \in W_s^{(0)}$, такие как $\Xi(x_0 + \varsigma_i | w_{0,s}^{(i)})$, где ς_i — случайные величины, $\mathbf{E}\varsigma_i = 0$, $\mathbf{E}\varsigma_i^2 < \varepsilon_\varsigma$. Гладкость функционала $\mathbf{\Delta}(x|\tau)$ достигается подстройкой параметров SVM-классификатора (τ и γ).

Изображения на рис. 3, *a*, *b* соответствуют «широкой» процедуре поиска $\mathbf{\Delta}_1(x|\tau)$, а на рис. 3, *c*, *d* — «узкой» процедуре поиска $\mathbf{\Delta}_2(x|\tau)$.

Структура функционала $\mathbf{\Delta}_i(x|\tau)$ для каждого этапа основывалась на различных настройках SVM-классификатора $CL_s(\cdot)$. В частности, для $\mathbf{\Delta}_1(x|\tau)$ использовалось сигмоидное ядро (гиперболический тангенс), а для $\mathbf{\Delta}_2(x|\tau)$ — классический RBF (Radial Basis Function).

Гиперпараметры SVM подбирались для каждого этапа индивидуально. В качестве алгоритма поиска применен эффективный метод Нелдера–Мида [18], который не использует производную от функционала $\mathbf{\Delta}_i(x|\tau)$. На каждой итерации поиска процедура Нелдера–Мида оценивает: насколько далеко центр x_k^+ поисковой рамки $\Xi(x_k^+)$ отстоит от точки экстремума целевого функционала $\mathbf{\Delta}_i(x|\tau)$. Экстремум функционала соответствует оценке величины x_0, а k — номер итерации. Процесс поиска останавливается по определенному правилу, момент останова соответствует точке x^+, которая объявляется оценкой для x_0. В процессе численных экспериментов среднее число поисковых итераций при определении величины x^+ составило величину равную 12, а среднее число вычислений значения целевого функционала — в диапазоне 30–40 (зависит от значения параметра p). В численных экспериментах, в качестве показателя точности оценивания координат экстремальной точки функционала $\mathbf{\Delta}_i(x|\tau)$, использовалась статистическая оценка величины

$$\mathbf{E}\|x^+ - x_0\|,$$

где x^+ — полученная в результате поиска оценка целевого параметра.

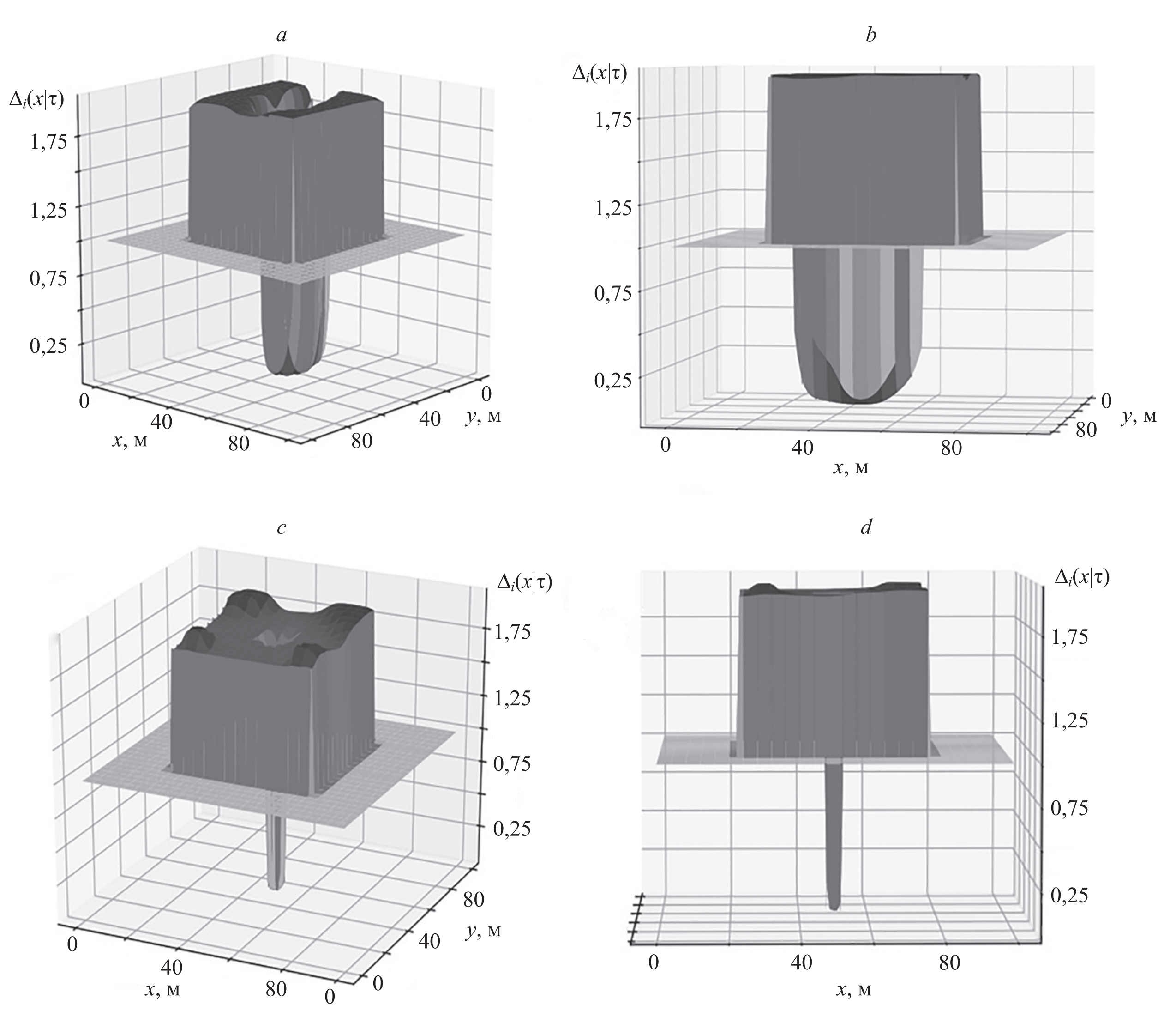


Рис. 3. Вид функционала $\mathbf{\Delta}_i(x|\tau)$ процедуры поиска для первого (*a*, *b*) и второго (*c*, *d*) этапов

Fig. 3. View of the search procedure functional $\mathbf{\Delta}_i(x|\tau)$ for the first (*a*, *b*) and second (*c*, *d*) steps

Для оценки $\mathbf{E}\|x^+ - x_0\|$ процедура поиска запускалась N раз, с различными начальными условиями и различными реализациями шумового искажения. При практической реализации алгоритма, согласно Теореме, останавливать процесс поиска можно в момент $i^+ = \underset{i}{\mathrm{ArgMin}}\{i|\rho_{\mathrm{CL}}(\chi[\Xi(x_i^+)]) \geq c_\alpha \neq \boldsymbol{\delta}\}$. В этом случае: $x^+ = x_{i^+}{}^+$. Кроме этого способа, в процессе численного исследования использовались и иные критерии останова. В процессе вычислительных экспериментов полагалось N = 5.

Результаты исследования алгоритма

Качество функционирования предложенного алгоритма исследовалось на размытых изображениях различного типа, соответствующих разным уровням помеховой обстановки. В том числе рассматривались как искусственные модели, так и реальные изображения.

Исследование на искусственных изображениях. Рассматривались две модели изображений:

1. «Острая» модель: радиус корреляции в центре 12, $z(x, y) = A_1[(1 + B_y)(1 + B_x)]^{-1} + \xi(x, y)$, где x, y — независимые переменные; z — ордината. $A_1 = 10$, $B_y = (y - (-50 + c_y)0{,}2)^2$, $B_x = (x - (-50 + c_x)0{,}2)^2$, где c_x, c_y — координаты центра искомого изображения, $c_x = 46$, $c_y = 48$.
2. «Плоская» модель: радиус корреляции в центре равен 18. $z(x, y) = A_1F(x, y) + A_2U(x, y) + \xi(x, y)$, где $\xi(x, y)$ — пространственно распределенный, коррелированный шум; $F(x, y) = \sin(2x)\sin(y)(x, y)^{-1}$; $U(x, y) = \sin(C_2 + x)\sin(C_1 + y)((C_2 + x)(C_1 + y))^{-1}$; $C_1 = 2{,}5$; $C_2 = 2{,}2$; $A_1 = 10$; $A_2 = 10$.

На рис. 4, *a* представлены изображения тестовой функции «плоская» модель, а на рис. 4, *b* — «острая» модель.

Функции отличаются значениями величин радиусов корреляции. Верхние фрагменты обоих частей содержат по четыре изображения, соответствующие различным вариантам аддитивного шума, который представлял собой гауссово поле $\xi(x, y)$ с $\mathbf{E}\xi(x, y) = 0$ и значениями дисперсии $\mathbf{E}\xi^2(x, y) \in \{0,1,2,4\}$, при пространственной корреляции на уровне 0,3.

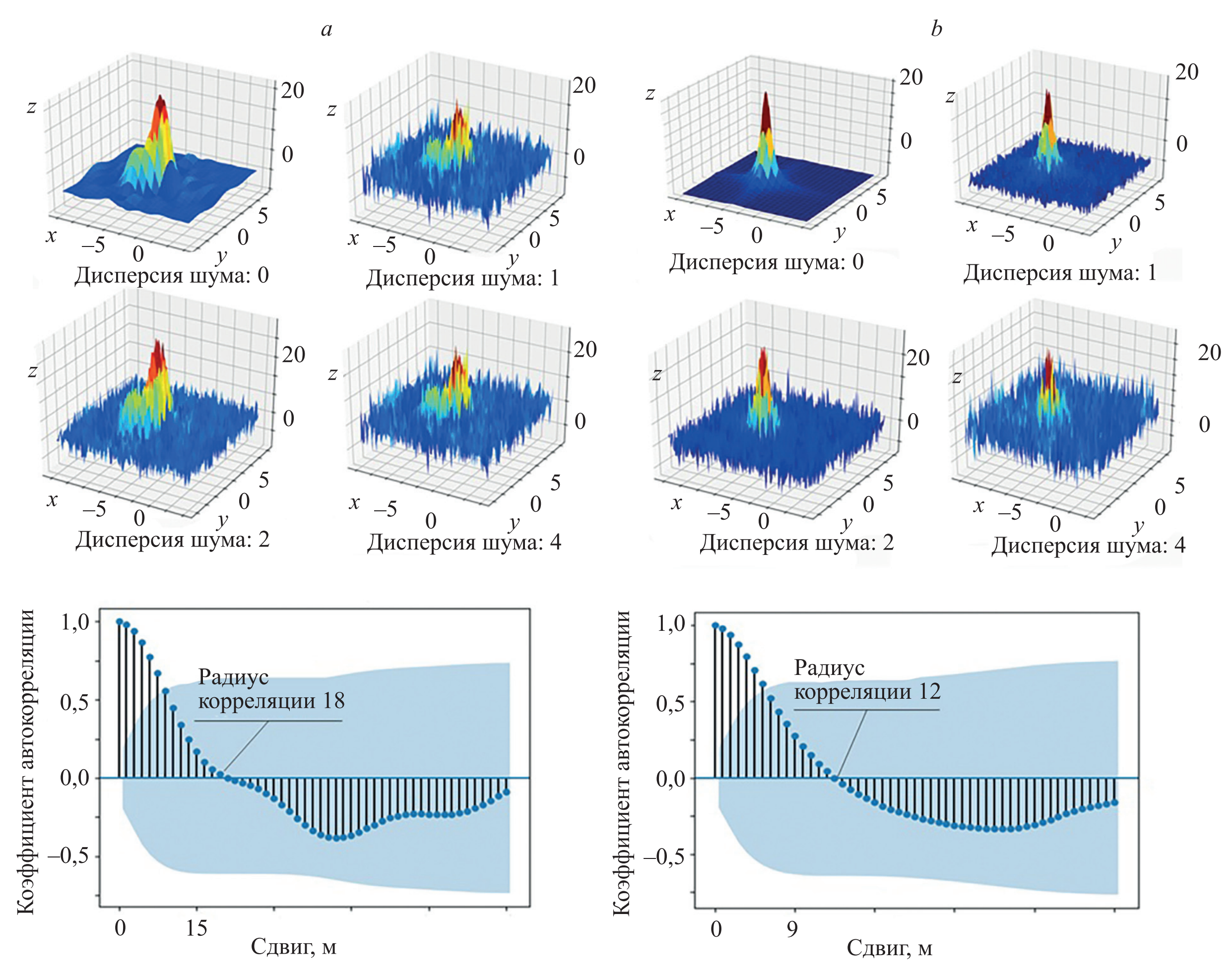


Рис. 4. Изображения тестовых функций с различными радиусами корреляции: «плоская» (*a*) и «острая» (*b*) модели

Fig. 4. Images of test functions with different correlation radii: "flat" (*a*) and "sharp" (*b*) models

На рис. 5 представлены результаты сравнения точности классического корреляционного метода и предложенного метода на «плоской» (рис. 5, *a*) и на «острой» (рис. 5, *b*) моделях, в зависимости от размера рамки (DLT — полуширина поисковой рамки, DLT = $|\Xi(\cdot)|2^{-1}$) при различных дисперсиях шумов. Рис. 5, *a*, *b* содержат по четыре изображения, соответствующие величинам $\mathbf{E}\xi^2(x, y) \in \{2,4,6,8\}$. Параметр точности на оси ординат: оценка величины $AE(x^+, x_0) = 100\mathbf{E}\|x^+ - x_0\|\cdot\|x_0\|^{-1}$. Из экспериментального материала следует, что с возрастанием величины $\mathbf{E}\xi^2(x, y)$ параметр $AE(x^+, x_0)$ увеличивается, т. е. точность плавно ухудшается. Однако существенно большее значение имеет размер поисковой рамки. Для рассмотренных моделей оптимальный размер рамки совпадает с размером «идеальной» рамки, при котором обеспечивается оптимальная точность. Для обеих моделей оптимальный размер рамки равен 24. Из данных эксперимента следует, что модель с меньшим радиусом корреляции («острая») обеспечила 25–30 % преимущество по точности. Важно, что классический корреляционный метод оценивания x_0 **существенно проигрывает** предложенному методу по точности. Например, при $|\Xi(\cdot)| = 24$ проигрыш по точности составляет около 100 %.

На рис. 6 представлена зависимость точности оценивания x_0 предложенным методом в виде зависимости параметра $AE(x^+, x_0)$ от полуширины поисковой рамки DLT, для случая «плоской» модели при различных величинах $\mathbf{E}\xi^2(x, y) \in \{1,2,4,6,8\}$. Для оптимального размера рамки ($|\Xi(\cdot)| = 24$) параметр $AE(x^+, x_0)$ — минимален, т. е. достигнута наилучшая точность. С другой стороны, что ожидаемо, увеличение параметра $\mathbf{E}\xi^2(x, y)$ ухудшает точность оценивания, но достаточно плавно. Таким образом, алгоритм робастен к шумам.

На рис. 7 представлена зависимость величины $AE(x^+, x_0)$ от степени зашумления изображения $\mathbf{E}\xi^2(x, y) \in \{1, …, 15\}$, для моделей разного радиуса корреляции («плоская»: серый график и «острая»: черный график). Радиусы корреляции в центре, соответственно, 12 и 18. Здесь $|\Xi(\cdot)| = 24$. Из графика следует, что модель с большим радиусом корреляции («плоская») проигрывает по точности модели с меньшим радиусом корреляции («острая»). Проигрыш примерно на уровне 25–30 %. При увеличении дисперсии шума точность ухудшается для обоих изображений, но деградация точности носит достаточно плавный характер, что свидетельствует о робастности алгоритма к аддитивному шуму.

На рис. 8 представлена зависимость точности от $\mathbf{E}\xi^2(x, y) \in \{1, …, 15\}$ для «широкой» модели с 95 % доверительными интервалами для величины $AE(x^+, x_0)$. Здесь $|\Xi(\cdot)| = 24$. Доверительные интервалы изображены вертикальными отрезками.

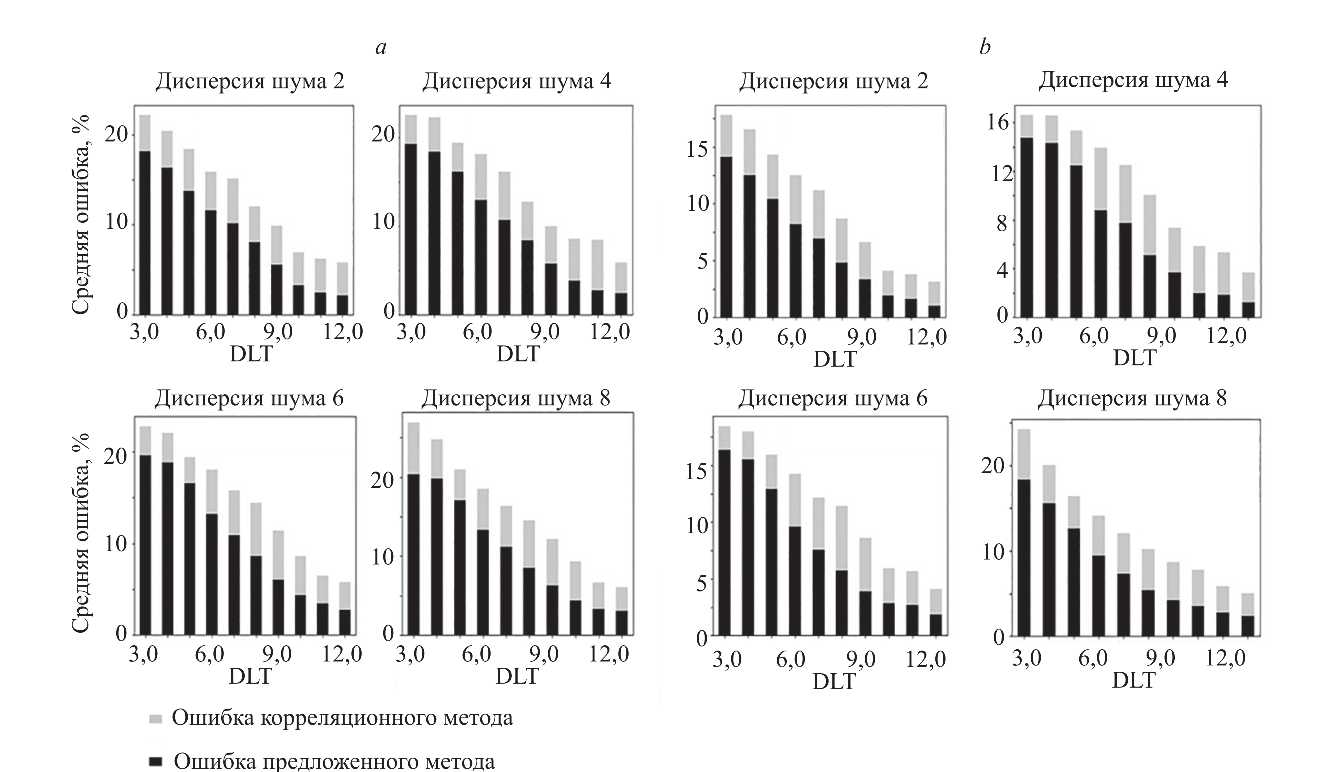


Рис. 5. Сравнения точности оценивания классического и предложенного методов для «плоской» (*a*) и «острой» (*b*) моделей

Fig. 5. Comparisons of estimation accuracy of classical and proposed methods for "flat" (*a*) and "sharp" (*b*) models

Для исследования робастности алгоритма к воздействию аддитивных шумов рассмотрена следующая функция:

$$\Phi(i+1) = (\mathbf{E}\|x^{+}_{i+1} - x_0\| / \mathbf{E}\|x^{+}_{i} - x_0\|)(\mathbf{E}\xi^2_{i+1}(x)/\mathbf{E}\xi^2_{i}(x))^{-1}$$

$$\forall i > 0\colon \mathbf{E}\xi^2_{i+1}(x) = \mathbf{E}\xi^2_{i}(x) + 1.$$

Поведение $\Phi(i+1)$, $i > 0$ наглядно демонстрирует темп ухудшения точности при увеличении дисперсии аддитивного шума. Согласно (1), для того чтобы алгоритм был R-робастен к аддитивному шуму, достаточно, чтобы $\forall i > 0$: $[(\mathbf{E}\xi^2_{i+1}(x) < R) \Rightarrow (\Phi(i+1) \leq 1)]$. В эксперименте использовалась «широкая» модель при $|\Xi(\cdot)| = 24$. Из рис. 9 следует, что $\Phi(i+1) < 1$, при увеличении $\mathbf{E}\xi^2(x)$ от 1 до 14, при этом $\Phi(i+1)$ возрастает достаточно плавно. Иначе говоря, предлагаемый алгоритм для данного изображения оказался R-робастным для $R = 14$.

На рис. 10 представлена зависимость величины $\mathrm{Diam}(\Upsilon)\|x_0\|^{-1}$ относительного диаметра байесовского доверительного множества Υ для величины x_0 при раз-

Рис. 6. Точность оценивания x_0 в зависимости от размера рамки и дисперсии шумов

Fig. 6. Estimation accuracy of x_0 as a function of frame size and noise variance

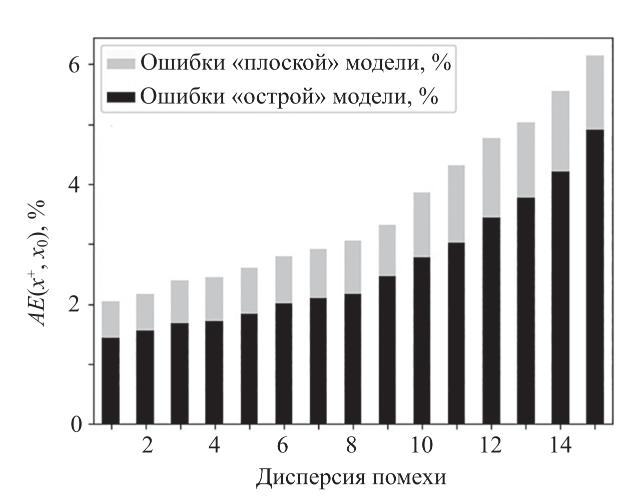


Рис. 7. Зависимость математического ожидания относительной ошибки от величины дисперсии помехи для двух изображений с различными радиусами корреляции

Fig. 7. The dependence of the relative error expectation on the noise variance for two images with different correlation radii

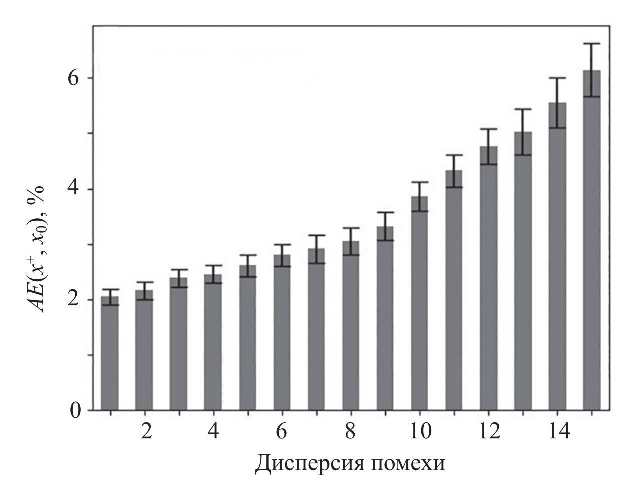


Рис. 8. Зависимость величины математического ожидания относительной ошибки от величины дисперсии помехи (с 95 % доверительными интервалами)

Fig. 8. Dependence of the relative error expectation on the interference variance (with 95 % confidence intervals)

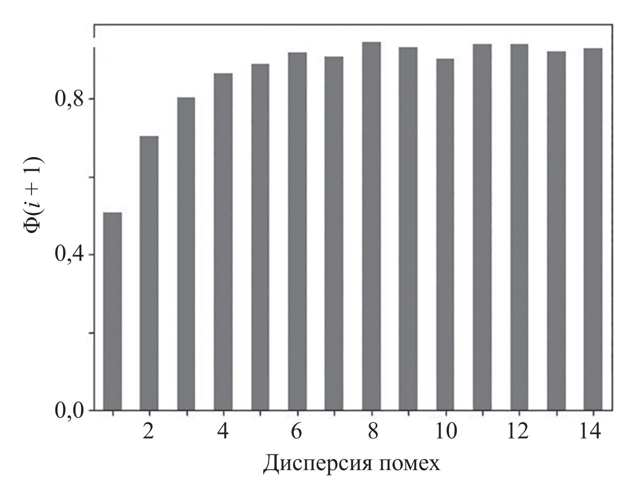


Рис. 9. Робастность алгоритма к влиянию аддитивного шума

Fig. 9. Robustness of the algorithm to the additive noise

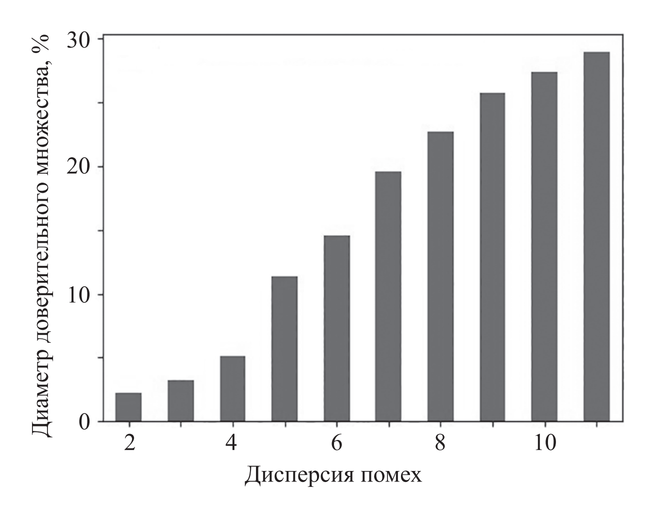


Рис. 10. Зависимость диаметра 95 % доверительного интервала для x_0 от дисперсии аддитивного шума

Fig. 10. Dependence of the diameter of the 95 % confidence interval for x_0 on the additive noise variance

личных дисперсиях аддитивных помех. Здесь DLT = 11, $\mathbf{P}(x_0 \in \Upsilon) \geq 0{,}95$, а численные исследования реализованы для «широкой» модели. Из приведенной зависимости следует, что величина $\mathrm{Diam}(\Upsilon)\|x_0\|^{-1}$ находится в практически приемлемых границах (не превышает 10 %) при сравнительно небольших шумах. А начиная с некоторого значения $\mathbf{E}\xi^2(x, y)$ (в конкретном примере, с величины дисперсии 5) резко возрастает. В этой связи имеет смысл использовать доверительное множество Υ для небольших дисперсий шумов. Практически, это вполне допустимое ограничение.

Исследование алгоритма на реальных изображениях

На рис. 11–14 приведены результаты работы алгоритма на четырех реальных сценах:

— оригинальные и зашумленные (размытые) сцены (рис. 11–14, *a*, *b*). Поступающие на обработку, зашумленные сцены соответствуют $\mathbf{E}\xi^2(x, y) = 7$, коэффициент пространственной корреляции шума 0,3. Поиск искомого изображения происходит на искаженной помехами сцене;

— автокорреляционные функции сцен по двум осям в точке x_0 для исходного и зашумленного изображений (рис. 11–14, *c*, *d*);

— результаты поиска целевой картинки на зашумленном изображении (рис. 11–14, *e*). По оси ординат отложена относительная ошибка оценивания $AE(x^+, x_0)$, в процентах, а по оси абсцисс — величина $\mathbf{E}\xi^2(x, y) \in \{1, \ldots, 10\}$. Приведены 95 % доверительные интервалы для $AE(x^+, x_0)$. В целом результаты работы алгоритма хорошие.

Для сцены «трактор на фоне песчаного грунта» (рис. 11) при $\mathbf{E}\xi^2(x, y) \leq 5$ имеем $AE(x^+, x_0) \leq 2$ %, обеспечение вероятности правильной классификации типа объекта 100 %.

Случай «автомобиль на фоне лиственного леса» (рис. 12) — самый сложный. Но и тут: для $\mathbf{E}\xi^2(x, y) \leq 2$ имеем $AE(x^+, x_0) \leq 5$ %. Вполне приемлемый практически результат.

Случаи на рис. 13 и 14 — сравнительно просты для предложенного алгоритма, и он справляется с ними довольно уверенно.

Для сцен, изображенных на рис. 11–14, и соответствующих им типов искомых изображений, множества $\Omega^{(0)}$ и $\Omega^{(1)}$ для этапа настройки (обучения) системы формировались по одной и той же схеме.

Шаг 1. Множество $\mathbf{\Omega}^{(0)} \cup \mathbf{\Omega}^{(1)}(|\mathbf{\Omega}^{(0)} \cup \mathbf{\Omega}^{(1)}| = 500)$ составляют зашумленные образцы сцен, и определено его разбиение на уровни $W_s^{(0)} \cup W_s^{(1)}$ масштаба s, в соответствии с масштабной сеткой $S^{(L)} = \{s_i | i = 1,\ldots, n_s\}$, $n_s = 5$, шаг $\upsilon_s = 5$ %.

Шаг 2. Каждый уровень масштаба $s \in S^{(L)}$ содержит такие изображения $\Omega_{s,\theta,l}^{(L)}(x_0) = \mathbf{B}[\mathbf{f}^{(s,\theta,l)}(x_0)]$, что

$$\forall \mathbf{f}^{(s,\theta,l)}(x_0) \exists (l \in \mathbf{L}, \theta \in \mathbf{\Theta}^{(L)}, k \in N_{s,\theta,l}^{(L)}):$$
$$:[\omega_k(x_0)(x) = \mathbf{f}^{(s,\theta,l)}(x_0)(x) + \xi_k^{(L)}(x)],$$

где $\mathbf{f}^{(s,\theta,l)}(x_0)(x) = A(\theta, 0, s)\mathbf{f}_{sp}^{(l)}(x_0)(x)$, $\Omega_{s,\theta,l}^{(L)}(x_0) = \{\omega_k(x_0)(x) | k \in N_{s,\theta,l}^{(L)}, x \in C[\Omega_{s,\theta,l}^{(L)}(x_0)]\}$, $n_{s,\theta,l}^{(L)} = |\Omega_{s,\theta,l}^{(L)}(x_0)|$, $N_{s,\theta,l}^{(L)} = \{1, \ldots, n_{s,\theta,l}^{(L)}\}$.

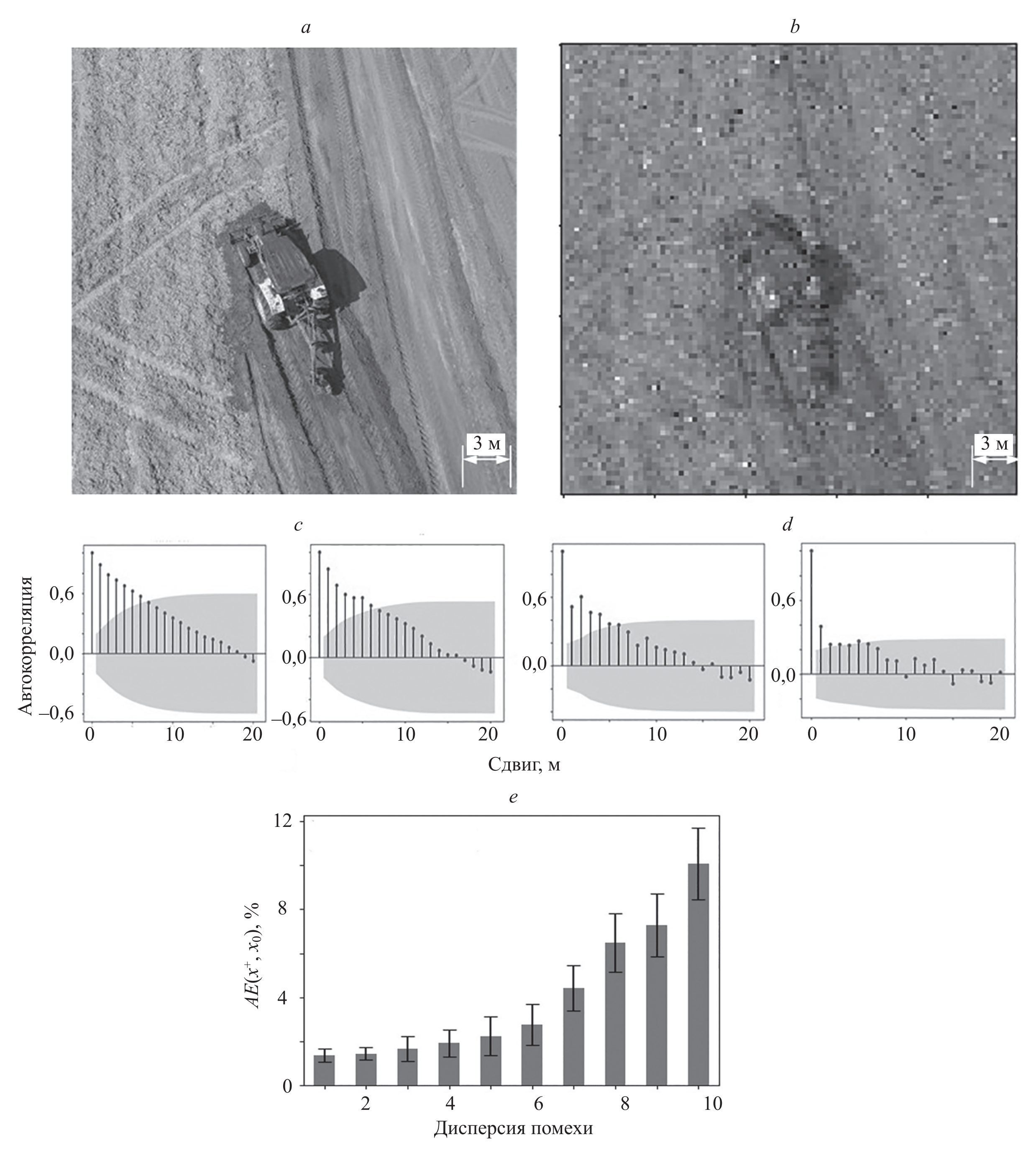


Рис. 11. Изображение «трактор на фоне песчаного грунта»: оригинальная (*a*) и зашумленная (размытая) (*b*) сцены; автокорреляционные функции (*c*, *d*) и результаты поиска целевой сцены на зашумленном изображении (*e*)

Fig. 11. The image "tractor on sandy ground": original (*a*) and noisy (blurred) (*b*) scenes; autocorrelation functions (*c*, *d*) and the result of searching for the target scene on the noisy image (*e*)

Ротационная сетка $\Theta^{(L)}$ такова, что $n_{\Theta^{(L)}} = 72$, а шаг углового перебора $\upsilon_\Theta = 5°$.

На этапе обучения системы, для оценивания обобщающей способности, использовалась стандартная схема скользящего контроля, иначе называемая CV (Cross Validation), в варианте LOO (leave-one-out). Для обеспечения этапа тестирования множество $\mathbf{\Omega}^{(0)}\bigcup\mathbf{\Omega}^{(1)}$ не используется. Напротив, формируются множества следующего типа:

$$\Omega^{(T)} = \{\Omega^{(T)}_{s,\theta,l}(x_0) | \Omega^{(T)}_{s,\theta,l}(x_0) = \mathbf{B}[\mathbf{f}^{(s,\theta,l)}(x_0)]\},$$

где

$$\forall \mathbf{f}^{(s,\theta,l)}(x_0) \exists (l \in \mathbf{L}, \theta \in \mathbf{\Theta}^{(T)}, s \in S^{(T)}, k \in N^{(L)}_{s,\theta,l}):$$

$$:\omega_k(x_0)(x) = \mathbf{f}^{(s,\theta,l)}(x_0)(x) + \xi^{(T)}_k(x).$$

$$\Omega^{(T)}_{s,\theta,l}(x_0) = \{\omega_k(x_0)(x) | k \in N^{(T)}_{s,\theta,l}, x \in C[\Omega^{(0)}_i(x_0)]\},$$

$$n^{(T)}_{s,\theta,l} = |\Omega^{(T)}_{s,\theta,l}(x_0)|, N^{(T)}_{s,\theta,l} = \{1, \ldots, n^{(T)}_{s,\theta,l}\}.$$

Принципиально важно то, что $\forall x,i: \xi^{(T)}_i(x) \neq \xi^{(T)}_i(x)$, $\Theta^{(T)} \neq \Theta^{(L)}$, а $S^{(T)} \neq S^{(L)}$. Неформально: помехи — различны, ротационные сетки — различны, масштабные сетки — различны. $|\Omega^{(T)}| = 500$.

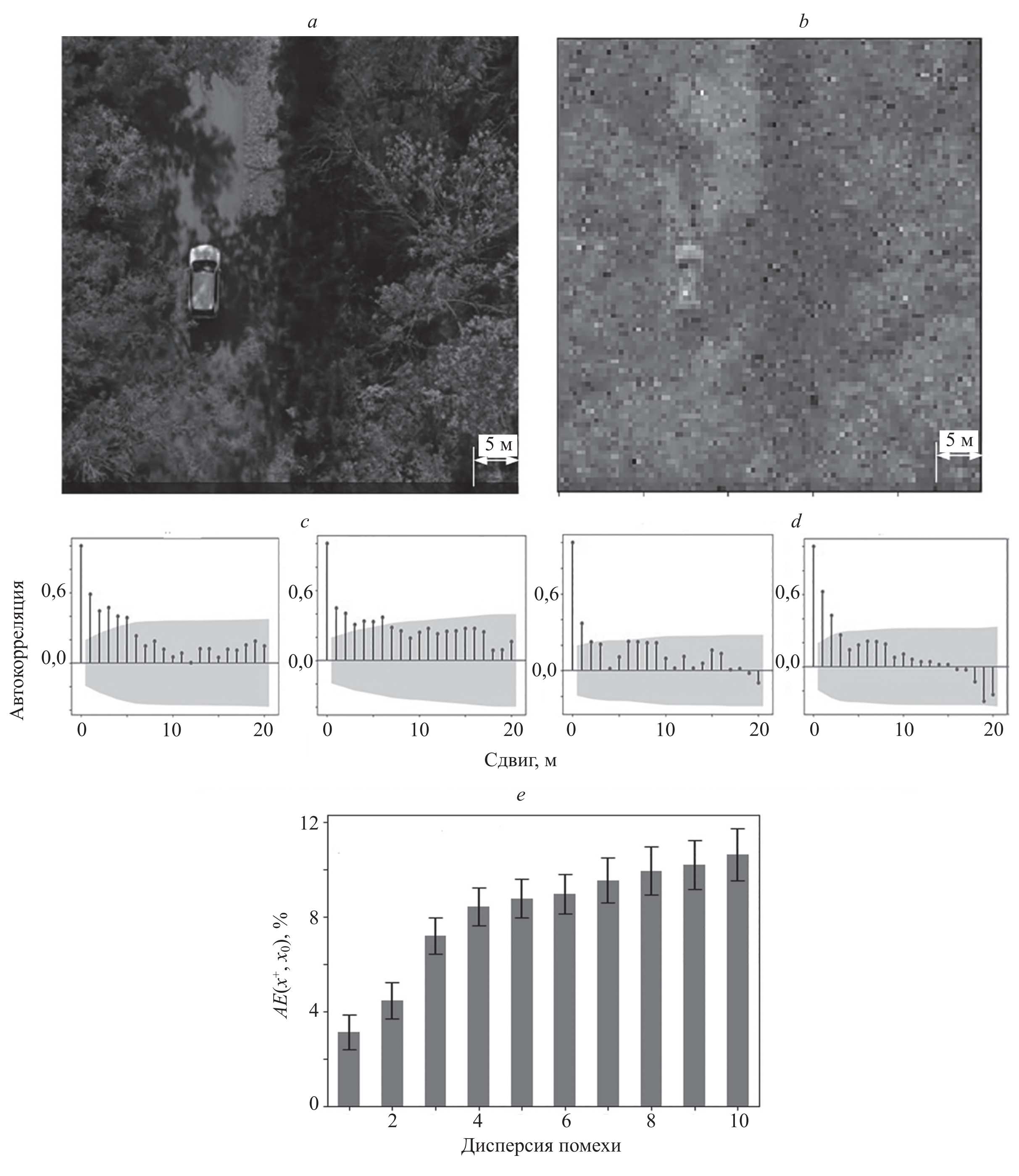


Рис. 12. Изображение «автомобиль на фоне лиственного леса»: оригинальная (*a*) и зашумленная (размытая) (*b*) сцены; автокорреляционные функции (*c*, *d*) и результаты поиска целевой сцены на зашумленном изображении (*e*)

Fig. 12. The image "car on the background of a deciduous forest": original (*a*) and noisy (blurred) (*b*) scenes; autocorrelation functions (*c*, *d*) and the result of searching for the target scene on the noisy image (*e*)

Все изображения, которые исследовались в процессе численных экспериментов, были **ε**-гладкими, для **ε** = 0,5. Методы SIFT, SURF, ORB и FAST оказались совершенно неработоспособны для изображения рассматриваемого типа.

Важный вопрос представляет собой выбор параметра p (число тестовых точек внутри поисковой рамки Ξ). Этот вопрос исследован на реальных изображениях нескольких типов. На рис. 15 приведены результаты этого исследования для случая, когда целевой класс представлял собой смесь из изображений двух типов (классов). Первый тип: «автомобиль на фоне асфальтового дорожного покрытия» и второй тип: «микроавтобус на фоне асфальтового покрытия». Использование «смесевого» класса, в качестве целевого, было сделано с целью демонстрации способности алгоритма устойчиво обнаруживать объекты, принадлежащие высоко вариабельному по составу классу, надежно отличая их от фона и достоверно оценивая их центр. Пример фрагмента корпуса, который использовался для обучения, приведен на рис. 15, *a*–*i*. Данная совокупность изображений приведена исключительно с целью де-

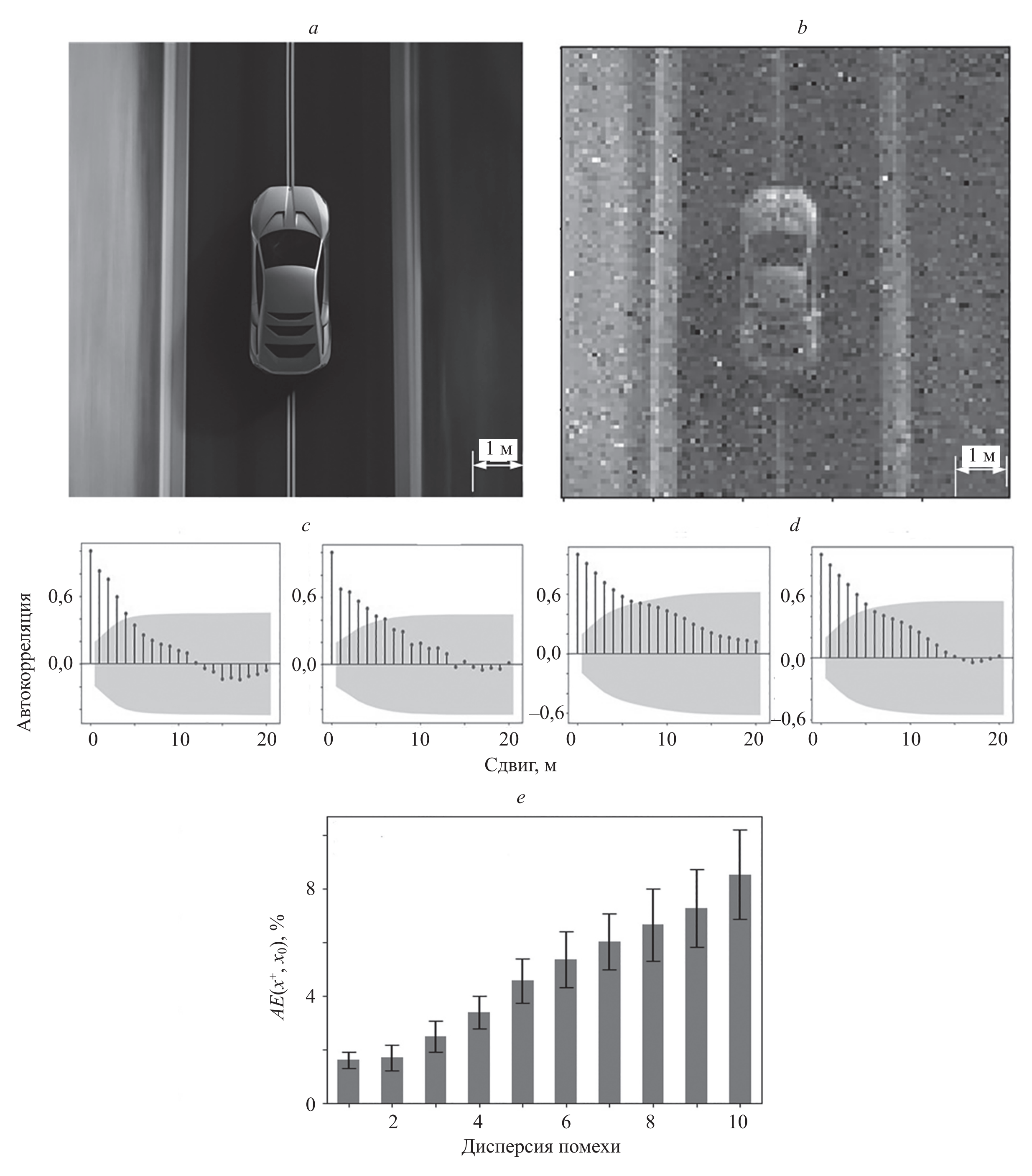


Рис. 13. Изображение «автомобиль на фоне асфальтового дорожного покрытия»: оригинальная (*a*) и зашумленная (размытая) (*b*) сцены; автокорреляционные функции (*c*, *d*) и результаты поиска целевой сцены на зашумленном изображении (*e*)

Fig. 13. The image "car on the background of the asphalt road surface": original (*a*) and noisy (blurred) (*b*) scenes; autocorrelation functions (*c*, *d*) and the result of searching for the target scene on the noisy image (*e*)

монстрации чрезвычайного низкого качества визуального материала, который способен эффективно обрабатывать предлагаемый алгоритм. Очевидно, что исходные изображения сильно размыты и зашумлены. Так как рассматривается «смесевой» класс, часть картинок относится к изображению легкового автомобиля (рис. 15, *a*–*c*, *g*, *h*), а часть — к изображению микроавтобуса (рис. 15, *d*–*f*, *i*). На рис. 15, *j* представлен график зависимости показателя точности (параметр $AE(x^+, x_0)$, с 95 % доверительными интервалами), оценивания параметра x_0 в зависимости от значения параметра *p*. На графике доверительные интервалы представлены вертикальными черными отрезками. Из графика следует, что наилучшее оценивание достигается при значении $p = 1936$. Но этому случаю соответствуют и наибольшие вычислительные затраты. По мере уменьшения величины p, точность плавно ухудшается, но даже для случая $p = 176$ (меньше «оптимальной» почти в 11 раз), величина $AE(x^+, x_0) \leq 3{,}3$ %, что вполне приемлемо с практической точки зрения.

Выполнено исследование влияния на устойчивость и точность решения поисковой задачи фактора рас-

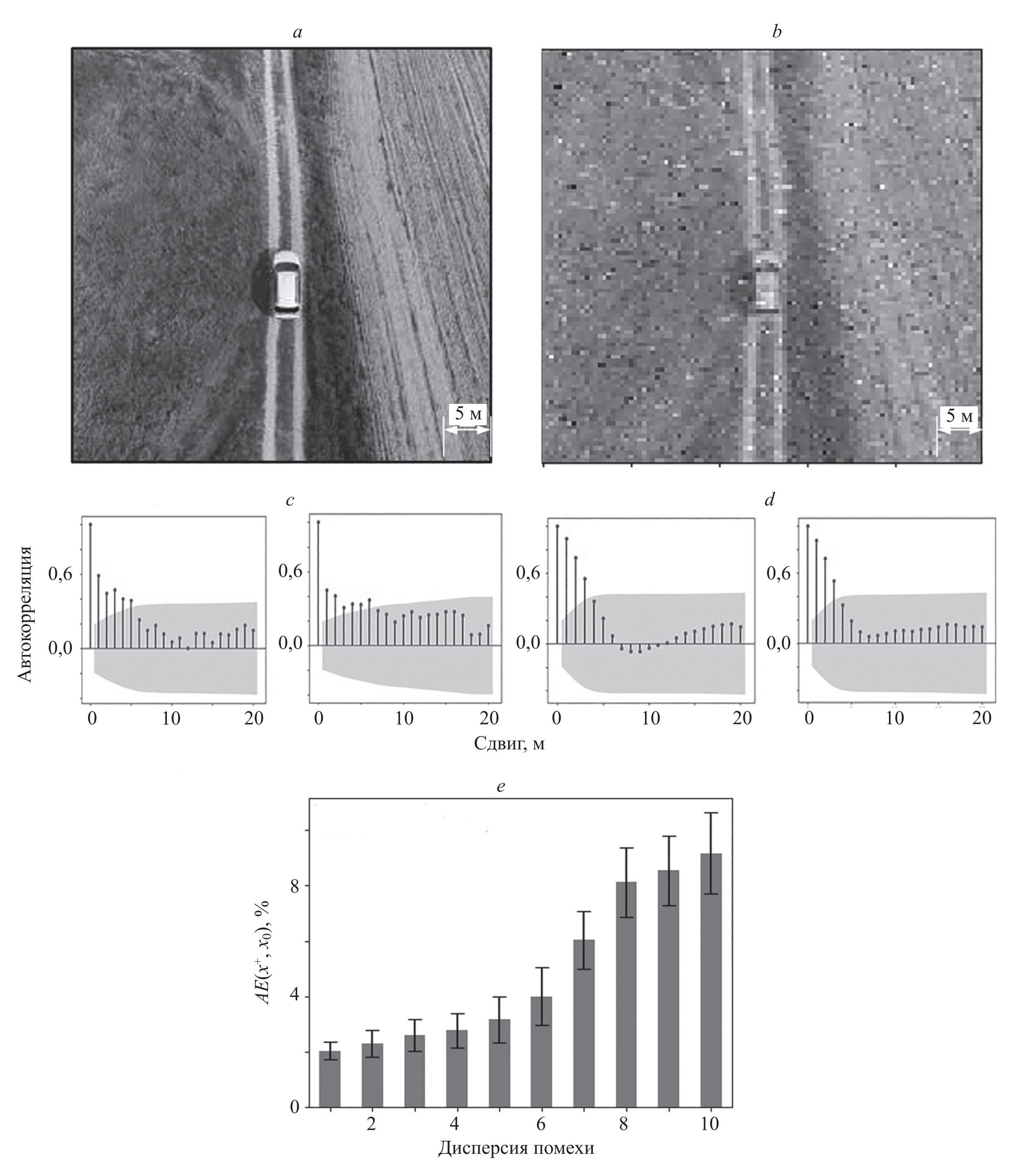
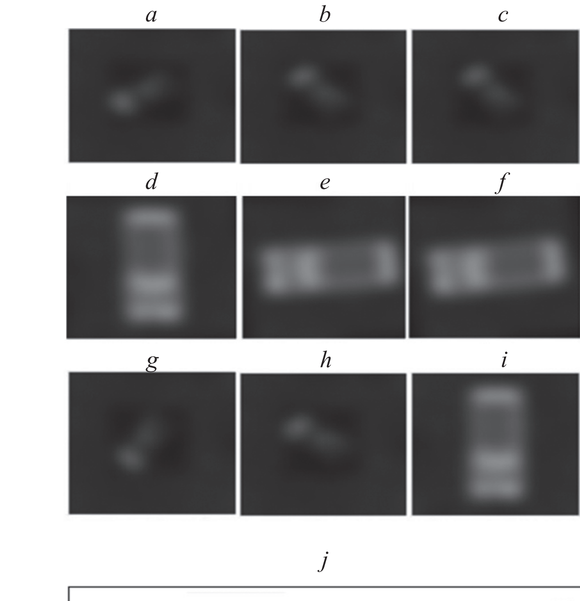


Рис. 14. Изображение «автомобиль на фоне грунтового дорожного покрытия»: оригинальная (*a*) и зашумленная (размытая) (*b*) сцены; автокорреляционные функции (*c*, *d*) и результаты поиска целевой сцены на зашумленном изображении (*e*)

Fig. 14. The image "car on the background of the unpaved road surface": original (*a*) and noisy (blurred) (*b*) scenes; autocorrelation functions (*c*, *d*) and the result of searching for the target scene on the noisy image (*e*)

согласованности масштабных сеток $S^{(T)}$ и $S^{(L)}$. Этот случай критически важен, так как точное совпадение масштабов изображения, на котором система была обучена, с масштабом $s^{(curr)}$ текущего изображения, на практике реализуется с очень низкой вероятностью. Есть все основания считать, что $s^{(curr)} \notin S^{(L)}$.

На рис. 16 приведены результаты исследования влияния несовпадения масштабов для ситуации, когда искомый класс состоял из смеси изображений двух типов (автомобиль и микроавтобус). Для тестирования принято: $S^{(T)} \neq S^{(L)}$, причем: $S^{(L)} = \{0, 8, 16, 20, 25\}$, а $S^{(T)} = \{0, 1, \ldots, 45\}$, т. е. $|S^{(T)} \bigcap S^{(L)}| = 5$. Устойчивость поисковой системы характеризуется зависимостью показателя «вероятность правильного обнаружения объекта» (P_{CD}) от значения параметра масштаба s (рис. 16, *a*). Точность поисковой системы характеризуется параметром средней ошибки оценивания координат центра искомого изображения (параметр ($AE(x^+, x_0)$) в зависимости от значения параметра масштаба s (рис. 16, *b*). Горизонтальная линия ограничива-



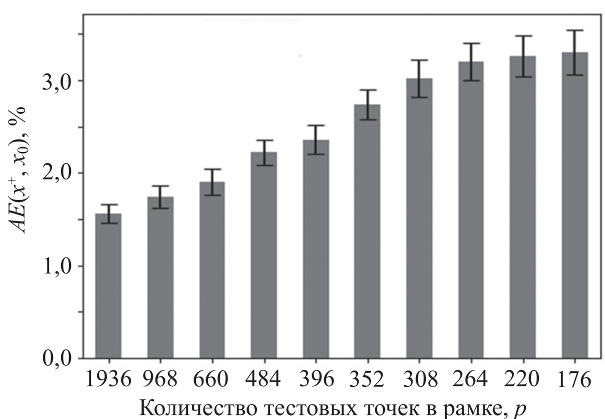


Рис. 15. Выбор количества тестовых точек внутри поисковой рамки (p): фрагмент корпуса, на котором производилось тестирование (a–i); величины $AE(x^+, x_0)$ с 95 % доверительными интервалами для различных значений параметра p (j)

Fig. 15. Selection of the number of test points within the search frame (p): fragment of the corpus on which testing was performed (a–i); average errors of the algorithm, with 95 % confidence intervals, for different values of parameter p (j)

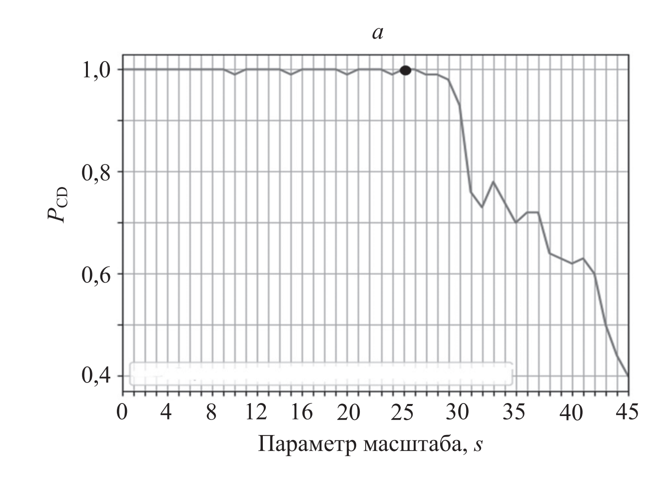




Рис. 16. Зависимости эффективности поиска от значения параметра масштаба s: вероятности правильного обнаружения объекта (P_{CD}) (a); величины $AE(x^+, x_0)$ (b)

Fig. 16. Dependence of search efficiency on scale mismatch: dependence of the probability of correct object detection on the value of the scale parameter s (a); dependence of $AE(x^+, x_0)$ on the value of the scale parameter s (b)

ет уровень 5 %, черные точки на рис. 16, b обозначают элементы $S^{(L)}$, черная точка на рис. 16, a — граничное значение из $S^{(L)}$ равное 25. Из рис. 16 следует, что система поиска сохраняла приемлемую точность $AE(x^+, x_0) \leq 4$ % (рис. 16, b) и устойчивость (рис. 16, a) вплоть до значения $s = 29$. Хотя «ближайшее» к этой точке значение масштаба из $S^{(L)}$ равнялось лишь 25. Причем зависимости на рис. 16 вполне согласованы между собой: на интервале (0, 25) система функционировала ожидаемо удовлетворительно $P_{CD} \geq 0{,}99$, $AE(x^+, x_0) \leq 4$ %, а на интервале (26, 29) продолжала оставаться работоспособной, обеспечивая $P_{CD} \geq 0{,}99$, $AE(x^+, x_0) \leq 5$ %. Быстрая деградация системы поиска началась примерно от значения $s = 30$, а уже при $s = 35$ было получено $P_{CD} = 0{,}7$, что неприемлемо практически. Таким образом, для обеспечения устойчивости алгоритма к фактору рассогласования масштабов, на интервале (0, 25), оказалось достаточно иметь всего пять опорных точек, составляющих множество $S^{(L)}$, в которых происходит машинное обучение системы.

Вывод: обученная (настроенная) описанным образом система сохраняет работоспособность для масштабов $S^{(T)} \backslash S^{(L)}$, обеспечивая $AE(x^+, x_0) \leq 5$ % при $P_{CD} \geq 0{,}99$.

Заключение

Метод, предложенный в настоящей работе в первую очередь предназначен для работы с изображениями низкого качества и малого размера, с плохо выраженными контурами. Численные исследования подтвердили его практическую работоспособность, а также состоятельность основных теоретических предпосылок. Для реализации метода на мобильной платформе потребуется перевод алгоритмов под архитектуру последовательного выполнения команд и в среде параллельных вычислений CUDA.

Литература

1. Lowe D.G. Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image. Patent US 6711293B1. 2004.
2. Bay H., Ess A., Tuytelaars T., Van Gool L. Speeded-Up Robust Features (SURF) // Computer Vision and Image Understanding. 2008. V. 110. N 3. P. 346–359. doi: 10.1016/j.cviu.2007.09.014
3. Rublee E., Rabaud V., Konolige K., Bradski G. ORB: An efficient alternative to SIFT or SURF // Proc. 2011 International Conference on Computer Vision (ICCV). 2011. P. 2564–2571. doi: 10.1109/ICCV.2011.6126544
4. Rosten E., Drummond T. Fusing points and lines for high performance tracking // Proc. 10th IEEE International Conference on Computer Vision (ICCV). 2005. V. 2. P. 1508–1515. doi: 10.1109/ICCV.2005.104
5. Vijaya Kumar B.V.K., Mahalanobis A., Juday R.D. Correlation Pattern Recognition. Cambridge University Press, 2005. 390 p. doi: 10.1017/CBO9780511541087
6. Тарасенко В.П., Тимофеев А.В. Доверительное оценивание точности совмещения в корреляционно-экстремальных системах // Автометрия. 1990. № 4. P. 106–111.
7. Kober V., Campos J. Accuracy of location measurement of a noisy target in a nonoverlapping background // Journal of the Optical Society of America A. 1996. V. 13. N 8. P. 1653–1666. doi: 10.1364/JOSAA.13.001653
8. Kim H.Y., De Araújo S.A. Grayscale template-matching invariant to rotation, scale, translation, brightness and contrast // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2007. V. 4872. P. 100–113. doi: 10.1007/978-3-540-77129-6_13
9. Korman S., Reichman D., Tsur G., Avidan S. FAsT-Match: Fast Affine Template Matching // Proc. 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2013. P. 2331–2338. doi: 10.1109/CVPR.2013.302
10. Rüping S. Robust Probabilistic Calibration // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2006. V. 4212. P. 743–750. doi: 10.1007/11871842_75
11. Timofeev A.V. Non-asymptotic sequential confidence regions with fixed sizes for the multivariate nonlinear parameters of regression // Statistical Methodology. 2009. V. 6. N 5. P. 513–526. doi: 10.1016/j.stamet.2009.05.002
12. Timofeev A.V. The guaranteed estimation of the Lipschitz classifier accuracy: confidence set approach // Journal of the Korean Statistical Society. 2012. V. 41. N 1. P. 105–114. doi: 10.1016/j.jkss.2011.07.005
13. Timofeev A.V. Nonasymptotic confidence sets of prescribed dimensions for parameters of nonlinear regressions // Automation and Remote Control. 2009. V. 70. N 2. P. 233–243. doi: 10.1134/S0005117909020052
14. Cristianini N., Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, 2000. XIII, 197 p. doi: 10.1017/CBO9780511801389
15. Mason L., Baxter J., Bartlett P., Frean M. Boosting algorithms as gradient descent // Advances in Neural Information Processing Systems. 2000. P. 512–518.
16. Friedman J.H. Greedy function approximation: a gradient boosting machine // Annals of Statistics. 2001. V. 29. N 5. P. 1189–1232. doi: 10.1214/aos/1013203451
17. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA. 2016. P. 785–794. doi: 10.1145/2939672.2939785
18. Nelder J.A. Mead R. A simplex method for function minimization // The Computer Journal. 1965. V. 7. N 4. P. 308–313. doi: 10.1093/comjnl/7.4.308

References

1. Lowe D.G. Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image. *Patent US 6711293B1*, 2004.
2. Bay H., Ess A., Tuytelaars T., Van Gool L. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 2008, vol. 110, no. 3, pp. 346–359. doi: 10.1016/j.cviu.2007.09.014
3. Rublee E., Rabaud V., Konolige K., Bradski G. ORB: An efficient alternative to SIFT or SURF. *Proc. 2011 International Conference on Computer Vision (ICCV)*, 2011, pp. 2564–2571. doi: 10.1109/ICCV.2011.6126544
4. Rosten E., Drummond T. Fusing points and lines for high performance tracking. *Proc. 10th IEEE International Conference on Computer Vision (ICCV)*, 2005, vol. 2, pp. 1508–1515. doi: 10.1109/ICCV.2005.104
5. Vijaya Kumar B.V.K., Mahalanobis A., Juday R.D. *Correlation Pattern Recognition*. Cambridge University Press, 2005, 390 p. doi: 10.1017/CBO9780511541087
6. Tarasenko V.P., Timofeev A.V. Confidence estimation of registration precision in correlation-extreme systems. *Avtometriya*, 1990, no. 4, pp. 106–111. (in Russian)
7. Kober V., Campos J. Accuracy of location measurement of a noisy target in a nonoverlapping background. *Journal of the Optical Society of America A*, 1996, vol. 13, no. 8, pp. 1653–1666. doi: 10.1364/JOSAA.13.001653
8. Kim H.Y., De Araújo S.A. Grayscale template-matching invariant to rotation, scale, translation, brightness and contrast. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007, vol. 4872, pp. 100–113. doi: 10.1007/978-3-540-77129-6_13
9. Korman S., Reichman D., Tsur G., Avidan S. FAsT-Match: Fast Affine Template Matching. *Proc. 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 2331–2338. doi: 10.1109/CVPR.2013.302
10. Rüping S. Robust Probabilistic Calibration. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, vol. 4212, pp. 743–750. doi: 10.1007/11871842_75
11. Timofeev A.V. Non-asymptotic sequential confidence regions with fixed sizes for the multivariate nonlinear parameters of regression. *Statistical Methodology*, 2009, vol. 6, no. 5, pp. 513–526. doi: 10.1016/j.stamet.2009.05.002
12. Timofeev A.V. The guaranteed estimation of the Lipschitz classifier accuracy: confidence set approach. *Journal of the Korean Statistical Society*, 2012, vol. 41, no. 1, pp. 105–114. doi: 10.1016/j.jkss.2011.07.005
13. Timofeev A.V. Nonasymptotic confidence sets of prescribed dimensions for parameters of nonlinear regressions. *Automation and Remote Control*, 2009, vol. 70, no. 2, pp. 233–243. doi: 10.1134/S0005117909020052
14. Cristianini N., Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000, XIII, 197 p. doi: 10.1017/CBO9780511801389
15. Mason L., Baxter J., Bartlett P., Frean M. Boosting algorithms as gradient descent. *Advances in Neural Information Processing Systems*, 2000, pp. 512–518.
16. Friedman J.H. Greedy function approximation: a gradient boosting machine. *Annals of Statistic*, 2001, vol. 29, no. 5, pp. 1189–1232. doi: 10.1214/aos/1013203451
17. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, 2016, pp. 785–794. doi: 10.1145/2939672.2939785
18. Nelder J.A. Mead R. A simplex method for function minimization. *The Computer Journal*, 1965, vol. 7, no. 4, pp. 308–313. doi: 10.1093/comjnl/7.4.308

Автор

Тимофеев Андрей Владимирович — доктор технических наук, научный директор, ТОО «Эквалайзум», Астана, 010000, Казахстан, SC 56689367600, https://orcid.org/0000-0001-7212-5230, timofeev.andrey@gmail.com

Author

Andrey V. Timofeev — D.Sc., Chief Scientific Officer, LLP EqualiZoom, Astana, 010000, Republic of Kazakhstan, SC 56689367600, https://orcid.org/0000-0001-7212-5230, timofeev.andrey@gmail.com

Статья поступила в редакцию 21.01.2021
Одобрена после рецензирования 11.02.2021
Принята к печати 16.03.2021

Received 21.01.2021
Approved after reviewing 11.02.2021
Accepted 16.03.2021

